Conversational Information Seeking

Other titles in Foundations and Trends® in Information Retrieval

Quantum-Inspired Neural Language Representation, Matching and Understanding
Peng Zhang, Hui Gao, Jing Zhang and Dawei Song
ISBN: 978-1-63828-204-4

Pre-training Methods in Information Retrieval
Yixing Fan, Xiaohui Xie, Yinqiong Cai, Jia Chen, Xinyu Ma, Xiangsheng Li, Ruqing Zhang and Jiafeng Guo
ISBN: 978-1-63828-062-0

Fairness in Information Access Systems
Michael D. Ekstrand, Anubrata Das, Robin Burke and Fernando Diaz
ISBN: 978-1-63828-040-8

Deep Learning for Dialogue Systems: Chit-Chat and Beyond
Rui Yan, Juntao Li and Zhou Yu
ISBN: 978-1-63828-022-4

Search Interface Design and Evaluation
Chang Liu, Ying-Hsang Liu, Jingjing Liu and Ralf Bierig
ISBN: 978-1-68083-922-7

Conversational Information Seeking

An Introduction to Conversational Search, Recommendation, and Question Answering

Hamed Zamani
University of Massachusetts Amherst
zamani@cs.umass.edu

Johanne R. Trippas
RMIT University
j.trippas@rmit.edu.au

Jeff Dalton
University of Edinburgh
jeff.dalton@ed.ac.uk

Filip Radlinski
Google Research
filiprad@google.com

the essence of knowledge
Boston — Delft

Foundations and Trends® in Information Retrieval

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

H. Zamani *et al.*. *Conversational Information Seeking*. Foundations and Trends® in Information Retrieval, vol. 17, no. 3-4, pp. 244–456, 2023.

ISBN: 978-1-63828-201-3

Foundations and Trends® in Information Retrieval
Volume 17, Issue 3-4, 2023

Editorial Scope

Topics

Foundations and Trends® in Information Retrieval publishes survey and tutorial articles in the following topics:

- Applications of IR
- Architectures for IR
- Collaborative filtering and recommender systems
- Cross-lingual and multilingual IR
- Distributed IR and federated search
- Evaluation issues and test collections for IR
- Formal models and language models for IR
- IR on mobile platforms
- Indexing and retrieval of structured documents
- Information categorization and clustering
- Information extraction
- Information filtering and routing
- Metasearch, rank aggregation and data fusion
- Natural language processing for IR
- Performance issues for IR systems, including algorithms, data structures, optimization techniques, and scalability
- Question answering
- Summarization of single documents, multiple documents, and corpora
- Text mining
- Topic detection and tracking
- Usability, interactivity, and visualization issues in IR
- User modelling and user studies for IR
- Web search

Information for Librarians

Foundations and Trends® in Information Retrieval, 2023, Volume 17, 5 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

Contents

Conversational Information Seeking

Hamed Zamani[1], Johanne R. Trippas[2], Jeff Dalton[3] and Filip Radlinski[4]

[1] *University of Massachusetts Amherst, USA; zamani@cs.umass.edu*
[2] *RMIT University, Australia; j.trippas@rmit.edu.au*
[3] *University of Edinburgh, UK; jeff.dalton@ed.ac.uk*
[4] *Google Research, UK; filiprad@google.com*

ABSTRACT

Conversational information seeking (CIS) is concerned with a sequence of interactions between one or more users and an information system. Interactions in CIS are primarily based on natural language dialogue, while they may include other types of interactions, such as click, touch, and body gestures. This monograph provides a thorough overview of CIS definitions, applications, interactions, interfaces, design, implementation, and evaluation. This monograph views CIS applications as including conversational search, conversational question answering, and conversational recommendation. Our aim is to provide an overview of past research related to CIS, introduce the current state-of-the-art in CIS, highlight the challenges still being faced in the community, and suggest future directions.

Hamed Zamani, Johanne R. Trippas, Jeff Dalton and Filip Radlinski (2023), "Conversational Information Seeking", Foundations and Trends® in Information Retrieval: Vol. 17, No. 3-4, pp 244–456. DOI: 10.1561/1500000081.

1

Introduction

1.1 Motivation

Over the years, information retrieval and search systems have become more *conversational*: For instance, techniques have been developed to support queries that refer indirectly to previous queries or previous results; to ask questions back to the user; to record and explicitly reference earlier statements made by the user; to interpret queries issued in fully natural language, and so forth. In fact, systems with multi-turn capabilities, natural language capabilities as well as robust long-term user modeling capabilities have been actively researched for decades. However, the last few years have seen a tremendous acceleration of this evolution.

This has been driven by a few factors. Foremost, progress in machine learning, specifically as applied to natural language understanding and spoken language understanding, has recently surged. Whereas the possibility of a conversational information seeking (CIS) system robustly understanding conversational input from a person was previously limited, it can now almost be taken for granted. In concert with this, consumer hardware that supports and encourages conversation has become common, raising awareness of — and the expectation of — conversational support in IR systems. From the research community, this has been accompanied by significant progress in defining more natural CIS tasks, metrics, challenges and benchmarks. This has allowed the field to expand rapidly. This monograph aims to summarize the current state of the art of conversational information seeking research, and provide an introduction to new researchers as well as a reference for established researchers in this area.

1.2 Guide to the Reader

The intended audience for this survey is computer science researchers in fields related to conversational information seeking, as well as students in this field. We do not assume an existing understanding of conversational systems. However, we do assume the reader is familiar with general concepts from information retrieval, such as indexing, querying and evaluation. As this monograph is not a technical presentation of recent machine learning algorithms, we also assume a basic understanding of machine learning and deep learning concepts and familiarity with key algorithms.

The reader will be provided with a summary of the open CIS problems that are currently attracting the most attention, and many promising current results and avenues of investigation. We will also provide an overview of applications attracting interest in the community, and the resources available for addressing these applications.

When discussing the structure of conversations we adopt terminology used in the speech and dialogue research community. The most basic unit is an *utterance* (analogous to a single query in retrieval). All contiguous utterances from a single speaker form a single *turn* (Traum and Heeman, 1996), with a conversation consisting of multiple turns from two or more participants. For the reader we note that somewhat confusingly, a commonly adopted definition in CIS publications defines a turn as the pair of a user turn and a system response turn (a user query and system answer).

The focus of this work differs from recent related surveys. We draw the reader's attention to the following most related examples. Gao *et al.* (2019) presented an overview focused on specific neural algorithmic solutions for question answering, task-oriented and chat agents. Freed (2021) also focused on the development of chatbots, often for customer support. Our focus is more on characterizing the problem space related to information seeking conversations and providing a broad overview of different problems, metrics and approaches. Moreover, the report from the third Strategic Workshop on Information Retrieval in Lorne (SWIRL 2018) (Culpepper *et al.*, 2018) provided a broader summary of important open challenges in information retrieval, where various challenges associated with CIS were ranked first. That document provides a briefer overview and reading list, more concretely aimed at summarizing open challenges. A more recent report from the Dagstuhl Seminar on Conversational Search (Anand *et al.*, 2020) reiterated these challenges in more detail. Beyond these, more focused recent relevant workshops include SCAI (Penha *et al.*, 2022), KaRS (Anelli *et al.*, 2022), Sim4IR (Balog *et al.*, 2022), Future Conversation (Spina *et al.*, 2021) and MuCAI (Hauptmann *et al.*, 2020) among others. Concurrent to this work, Gao *et al.* (2023) published a book draft on deep learning approaches for conversational information retrieval. This monograph provides a holistic overview of CIS systems, state-of-the-art CIS approaches, and future directions in CIS research. In contrast, Gao *et al.*'s book focuses specifically on deep learning solutions for various subtasks in conversational IR, therefore providing a complementary view to ours.

1.3 Scope

This monograph focuses on a particular class of conversational systems, namely those that exhibit key attributes of human conversation. We take a cue from Radlinski and Craswell (2017), who propose that a conversational system should incorporate mixed initiative (with both system and user able to take initiative at different times), memory (the ability to reference and incorporate past statements), system revealment (enabling the system to reveal its capabilities and corpus), user revealment (enabling the user to reveal and/or discover their information need), and set retrieval (considering utility over sets of complementary items). Here, we study approaches that exhibit at least some of these properties. In particular, we do not delve deeply into *dialogue systems* that restrict themselves largely to identifying slot/value pairs in back and forth exchanges between the system and user.

Additionally, we focus on *information seeking*, which refers to the process of acquiring information through conversation in order to satisfy the users' information needs. This implies that the conversation should exhibit a clear goal or assist the human user in completing a specific task through finding information. While significant progress has been recently made on chit-chat systems, with a primary goal of keeping users engaged in realistic conversational exchanges over a prolonged time (for more information, see Yan *et al.*, 2022), we do not attempt to cover such work in depth. Our focus thus aligns more with traditional search concepts such as the presence of an information need or user agenda that existed before they engaged with the CIS system, and which can be satisfied through a conversation.

On the other hand, we do not make a strong distinction between *search* and *recommendation* tasks. Rather, we cover both types of conversational information seeking interactions. We see these as strongly related tasks that are becoming more closely related as time passes. Indeed, we believe that the same task can often be characterized as either. For instance, a query "hotels in London" can be seen as either a search task (e.g. on a desktop interface, for a potential future tourist considering affordability in different areas) or a recommendation task (e.g. using a smart watch while standing in heavy rain in central London). Clearly device, interface and context play an important role in determining the best next conversational step.

Finally, we draw attention to three key aspects of CIS that, while having received significant attention, remain largely unsolved. First, the level of natural language understanding in conversational systems remains far from human-level, particularly over long sequences of exchanges. Even over adjacent conversational steps, question/answer interpretation remains challenging. Second, robust evaluation of conversational systems remains a critical research challenge: The highly personalized and adaptive nature of conversations makes test collection construction highly challenging. We will cover many of the common approaches and their limitations. Third, *conversation* is sometimes taken to imply voice or speech interactions. We do not make this assumption, recognizing that conversations can happen in many types of interfaces and modalities. We discuss research of conversations combining different types of interfaces and presentations in depth.

> Three particularly important aspects of CIS that are very active areas of research include obtaining human-level natural language understanding, robust evaluation of CIS systems, and moving beyond simple text and speech interactions.

There are a number of particularly important aspects of conversational information seeking that despite their importance are not covered in depth here, as they apply broadly across many non-conversational search and recommendation tasks. The first is the question of privacy. Clearly this is an essential aspect of all search tasks – and should be considered in depth in any practical system. We refer readers to Cooper (2008) and Zhang *et al.* (2016) as a starting point for privacy considerations as applied to logging and log analysis.

Similarly, we do not consider the type of information that a user may request or receive – including information that might be considered offensive or harmful. As this issue is not specific to conversational systems and is heavily studied; A detailed consideration of such information access is thus beyond our scope. We refer readers to Yenala *et al.* (2018) and Pradeep *et al.* (2021) as starting points of recent work in this space.

Along the same lines, fairness is an essential aspect for information seeking and recommendation tasks, yet largely beyond our scope. We note that this includes both fairness in terms of biases that may exist in recommendation to different groups (Ge *et al.*, 2021) as well as fairness when considering both consumers of recommendations as well as producers of items being recommended (Abdollahpouri *et al.*, 2020). We refer interested readers to Ekstrand *et al.* (2022) for a complete recent overview.

1.4 Applications

An alternative way to characterize the scope of this work could be in terms of the relevant *applications* that are addressed. Section 2 will focus on this formulation, starting with a brief introduction on conversational information seeking (Section 2.3). This includes a discussion of different modalities' (that is, text, speech, or multi-modal) impact on the seeking process, as for instance studied by Deldjoo *et al.* (2021). We then continue with the topic of conversational search and its various proposed definitions (Section 2.5), culminating with one that relates CIS to many other related settings (Anand *et al.*, 2020). Section 2.6 introduces conversational recommendation (Jannach *et al.*, 2021a) followed by conversational question answering in Section 2.7, where for instance Qu *et al.* (2019b) provide a powerful characterization of the relationships between these areas of study. We continue Section 2 by explaining how CIS applications can be used in different domains, and focus on e-commerce, enterprise, and health in Section 2.8. The section concludes with details on intelligent assistants with relation to CIS.

1.5 A High-Level Architecture for CIS Systems

To create a structure for the remainder of this work, we follow the general structure of most CIS systems. This choice guides the main body of this monograph: Each section in this part focuses on a core technological competency that is essential to a modern CIS system. In particular, a CIS system must first choose an interface (Section 1.5.1). It must then have an approach to maintain the state of a conversation (Section 1.5.2), and at each system turn determine the system's next utterance (Section 1.5.3). One particular challenge that is attracting attention is when the system should take initiative versus responding passively (Section 1.5.4).

> Key design considerations of a CIS system include its chosen interface, how it maintains conversational state, and how it selects the system's next utterance. One particular challenge for the latter is that of when the system should take initiative.

1.5.1 Conversational Interfaces and Result Presentation

Section 3 provides an overview of conversational interfaces. We begin with a historical perspective, where we explain differences between existing conversational interfaces such as spoken dialogue systems, voice user interfaces, live chat support, and chatbots. This overview illustrates the use of conversations within closely related CIS applications (McTear *et al.*, 2016). Next, research on result presentation through different mediums (desktop or small device) and modalities (text, voice, multi-modal) are discussed in Section 3.2, such as recent work by Kaushik *et al.* (2020). This overview emphasizes the difficulties with highly interactive result presentation and highlights research opportunities. Following this, Section 3.3 introduces different kinds of initiative in conversational systems, including system-initiative, mixed-initative, and user-initiative, for instance well-characterized by Zue and Glass (2000) and Wadhwa and Zamani (2021). This section aims to explain the different kinds of initiative, and the consequences on human-machine interactions. We finish the section with a discussion of conversational interfaces limitations including, for instance, limitations as experienced by visually impaired searchers (Gooda Sahib *et al.*, 2015).

1.5.2 Tracking and Understanding Flow

The focus of Section 4 is on the varying approaches that make it possible to follow conversational structure. We begin with an overview of how to represent a single turn, such as is done with Transformer models (Raffel *et al.*, 2020), and how turns are often classified into dialogue acts (Reddy *et al.*, 2019). Section 4.2 then looks at how the different turns of a conversation are usually tied together through state tracking and text resolution across turns. In particular, the structure of longer conversations is looked at in-depth in Section 4.3, although noting that existing models are often limited in their ability to capture long-distance conversational structure (Chiang *et al.*, 2020). We cover work that operates over long-term representation of CIS exchanges in Section 4.4, followed by recent work that attempts to model longer conversations in the final section, epitomized by work on selecting the right context for understanding each turn (Dinan *et al.*, 2019a).

1.5.3 Determining Next Utterances

The next step for a canonical conversational system is selecting or generating a relevant response in the conversational context. This is the focus of Section 5. We begin with an overview of the different types of responses, including short answers, long answers, and structured entities or attributes. The short answer section presents early Conversational QA (ConvQA) systems then discusses the transition to more recent Transformer architectures based on pre-trained language models. Section 5.1.5 then examines how ConvQA is performed over structured knowledge graphs including systems that use key-value networks (Saha *et al.*, 2018), generative approaches, and logical query representations (Plepi *et al.*, 2021). Following this, we discuss open retrieval from large text corpora as part of the QA process. In particular, Section 5.2 goes beyond short answer QA to approaches performing conversational passage retrieval from open text collections including multi-stage neural ranking, for instance recently considered by Lin *et al.* (2021b). We briefly discuss long answer generation approaches in Section 5.3 including both extractive and abstractive summarization methods. We conclude the section with conversational ranking of items in a recommendation context, including models that use multi-armed bandit approaches to trade-off between elicitation and item recommendation.

1.5.4 Initiative

Section 6 provides a detailed look at mixed-initiative interactions in CIS systems. We start with reviewing the main principles of developing mixed-initiative interactive systems, and describing different levels of mixed-initiative interactions in dialogue systems (Allen *et al.*, 1999; Horvitz, 1999). We briefly review system-initiative interactions with a focus on information seeking conversations, such as the work of Wadhwa and Zamani (2021), in Section 6.1. We then delve deeply into intent clarification as an example of important mixed-initiative interactions for CIS in Section 6.2. We introduce taxonomy of clarification and review models for generating and selecting clarifying questions, such as those by Aliannejadi *et al.* (2019) and Zamani *et al.* (2020a). In presenting the work, we include models that generate clarifying questions trained using maximum likelihood as well as clarification maximization through reinforcement learning. Additionally, Section 6.3 discusses preference elicitation and its relation with clarification, followed by mixed-initiative feedback (*i.e.*, getting feedback from or giving feedback to users via sub-dialogue initiation) in Section 6.4.

1.6 Evaluation

Beyond the details of how a CIS system functions, fair evaluation is key to assessing the strengths and weaknesses of the solutions developed. Section 7 looks at evaluation in CIS holistically. After considering possible ways of studying this broad space, this section breaks down evaluation by the setting that is evaluated. Specifically, offline evaluation is treated first, in Section 7.2. A variety of frequently used offline datasets are presented (such as Multi-WOZ (Budzianowski *et al.*, 2018)), and strengths and limitations are discussed including the use of simulators to produce more privacy-aware evaluations as well as the use of non-text datasets. Online evaluation is considered next, with Section 7.3 contrasting lab studies, crowdsourcing, and real-world evaluations. An example of these is where commercial systems may ask evaluation questions of their users (Park *et al.*, 2020). Finally, the metrics applied in these settings are covered in Section 7.4. While readers are referred to Liu *et al.* (2021a) for a full treatment, we present an overview of typical turn-level as well as end-to-end evaluation metrics.

1.7 Open Research Directions

Section 8 provides a brief summary of this monograph and discusses different open research directions. We collate the major themes discussed throughout this manuscript instead of presenting a detailed account of all possible future research problems. We highlight four key areas for future exploration. First, Section 8.2.1 covers challenges related to modeling and producing conversational interactions as a way to transfer information between user and system. Second, we highlight the importance of result presentation and its role in CIS research in Section 8.2.2. Third, we emphasise the importance of different CIS tasks in Section 8.2.3. Finally, Section 8.2.4 covers measures of success during the highly interactive CIS process and ultimate evaluation of CIS systems.

1.8 Further Resources

Beyond the main body of this work, Appendix A briefly presents a more holistic historical context for this monograph. This appendix mainly includes information about early research on interactive information retrieval, as well as on dialogue-based information retrieval, such as the I^3R (Croft and Thompson, 1987) and THOMAS (Oddy, 1977) systems (see Section A.1). We discuss approaches for theoretical modelling of interactive information retrieval systems, such as game theory-based models (Zhai, 2016) and economic models (Azzopardi, 2011) in Section A.2. We also include introductory information about existing literature on session search, such as the TREC Session Track, and evaluation methodologies for session search tasks (Carterette *et al.*, 2016) in Section A.3. Finally, we briefly cover exploratory search (White and Roth, 2009) and discuss its relationship to conversational information seeking in Section A.4, followed by a very brief overview of chit-chat and task-oriented dialogue systems in Section A.5. Newcomers to the field of information retrieval are highly encouraged to review this appendix to develop an understanding of where the core ideas behind CIS originated.

This monograph has been used in multiple tutorials on conversational information seeking at top-tier conferences, *e.g.*, at the SIGIR 2022 (Dalton *et al.*, 2022) and the Web Conference 2023 (Dalton *et al.*, 2023). The materials prepared for these tutorials, *e.g.*, presentation slides, interactive demos, and coding practices, are available at https://cis-tutorial.github.io/.

Acknowledgments

We would like to thank the researchers who contributed directly or indirectly to the field of conversational information seeking. We are also grateful to our colleagues, especially W. Bruce Croft, Reena Jana, and David Reitter, who gave us invaluable suggestions and helped us better position this monograph. We also thank the anonymous reviewers and the editors of Foundations and Trends® in Information Retrieval, especially Maarten de Rijke for his support throughout the process.

This work was supported in part by the Center for Intelligent Information Retrieval, in part by NSF grant number 2143434, in part by the Office of Naval Research contract number N000142212688, in part by an Alexa Prize grant from Amazon, and in part by an Engineering and Physical Sciences Research Council fellowship titled "Neural Conversational Information Seeking Assistant" grant number EP/V025708/1. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not reflect those of the sponsors or the authors' employers.

This research was partially completed when Johanne R. Trippas was affiliated with and funded by the University of Melbourne, and Jeff Dalton was affiliated with the University of Glasgow.

Appendices

A

Historical Context

In this appendix, we briefly provide a historical context to information retrieval and dialogue systems research related to conversational information seeking systems. Readers that are not familiar with early IR research are especially encouraged to read this appendix.

A.1 Interactive Information Retrieval Background

Conversational information seeking systems have roots in interactive information retrieval (IIR) research. The study of interaction has a long history in information retrieval research, starting in the 1960s (Kelly and Sugimoto, 2013). Much of the earlier research studied how users interacted with intermediaries (*e.g.*, librarians) during information seeking dialogues but this rapidly shifted to studying how users interacted with operational retrieval systems, including proposals for how to improve the interaction. Information retrieval systems based on this research were also implemented. Belkin (1980) studied the concept of Anomalous States of Knowledge (ASK) of users of IR systems and discussed the importance of multi-turn interactions to help user formulate their needs and help systems successfully retrieve relevant information (Belkin and Kwaundefinednik, 1986). Brooks and Belkin (1983) studied information seeking dialogues between a user and an intermediary and introduced a annotation coding scheme for discourse analysis of the dialogues.

Oddy (1977) developed an interactive information retrieval system with rule-based dialogue interactions in 1977, called THOMAS. Example snapshots of user interactions with THOMAS are presented in Figure A.1. As shown in the figure, THOMAS includes a few pre-defined interaction types. Even though THOMAS handles a sequence of interactions, it does not model users which is essential for IIR systems. Croft and Thompson (1987) closed this gap by proposing the I^3R system – the first IIR system with a user modeling component. I^3R uses a mixture of experts architecture. It assists users by accepting Boolean queries, typical text queries, and documents (query by examples). It enables users to provide explicit relevance feedback to the system. Example snapshots of user interactions with I^3R are presented in Figure A.2. Later on, Belkin *et al.* (1995) focused on user interactions with IIR systems and characterized information seeking strategies for interactive IR, offering users choices in a search session based on case-based reasoning. They defined a multi-dimensional space of information seeking strategies and applied their model to the MERIT system, a prototype IIR system that implements these multi-dimensional design principles.

Since the development of web search engines, research has mostly focused heavily on understanding user interaction with search engines based on an analysis of the search logs available to commercial search engine providers, *e.g.*, see Dumais *et al.* (2014), Buscher *et al.* (2009), Teevan *et al.* (2007), and Murray and Teevan (2007). Since then, explicit modeling of information seeking dialogues or conversations with the aim of improving the effectiveness of retrieval has not been a focus of research until recently. Among them, session search is perhaps the closest research area to CIS (see Section A.3).

A.2 Formal Modeling of IIR Systems

The proposition that IR systems are fundamentally interactive and should be evaluated from the users' perspective is not new (Kelly, 2009). This has been highlighted by many pioneers in the field since the 1960s (Cleverdon and Kean, 1968; Salton, 1970). However, today's

THOMAS, THE REFERENCE RETRIEVAL PROGRAM

Help can be obtained whenever the program has displayed the start symbol by typing '?' immediately after it.

Please give a short name for the search:
▶ Alv.Resp.

Start searching:
▶ pulmonary alveoli

Influence of fasting on blood gas tension, pH, and related values in dogs.; Pickrell *et al, Am J Vet Res, 34*, 805–8, Jun 73
1. J A Pickrell, 2. J L Mauderly, 3. B A Muggenburg, 4. U C Luft, 5. animal experiments, 6. animal feed, 7. arteries, 8. blood, 9. body temperature, 10. carbon dioxide, 11. dogs, 12. fasting, 13. hemoglobin, 14. hydrogen-ion concentration, 15. irrigation, 16. lung, 17. oxygen, 18. pulmonary alveoli, 19. respiration, 20. time factors
▶ ?

There can be three parts to your statement (all optional):

1. Your reaction to the reference just shown (if any).
   This must come first:
   "Yes" or "No"
2. A selection from the names (authors) or terms shown, by number. A "not" in the statement signifies rejection of all numbers that follow it.
3. New names or terms (terms preferably in quotes). The elements of the statement should be separated by commas.

Examples: 'posture', 'circulatory system'
Yes, not 11,12
No, 7,13,4
'heart rate'
Yes

Press enter key when you are ready to proceed ▶

Figure A.1: Snapshots from the THOMAS system (Oddy, 1977).

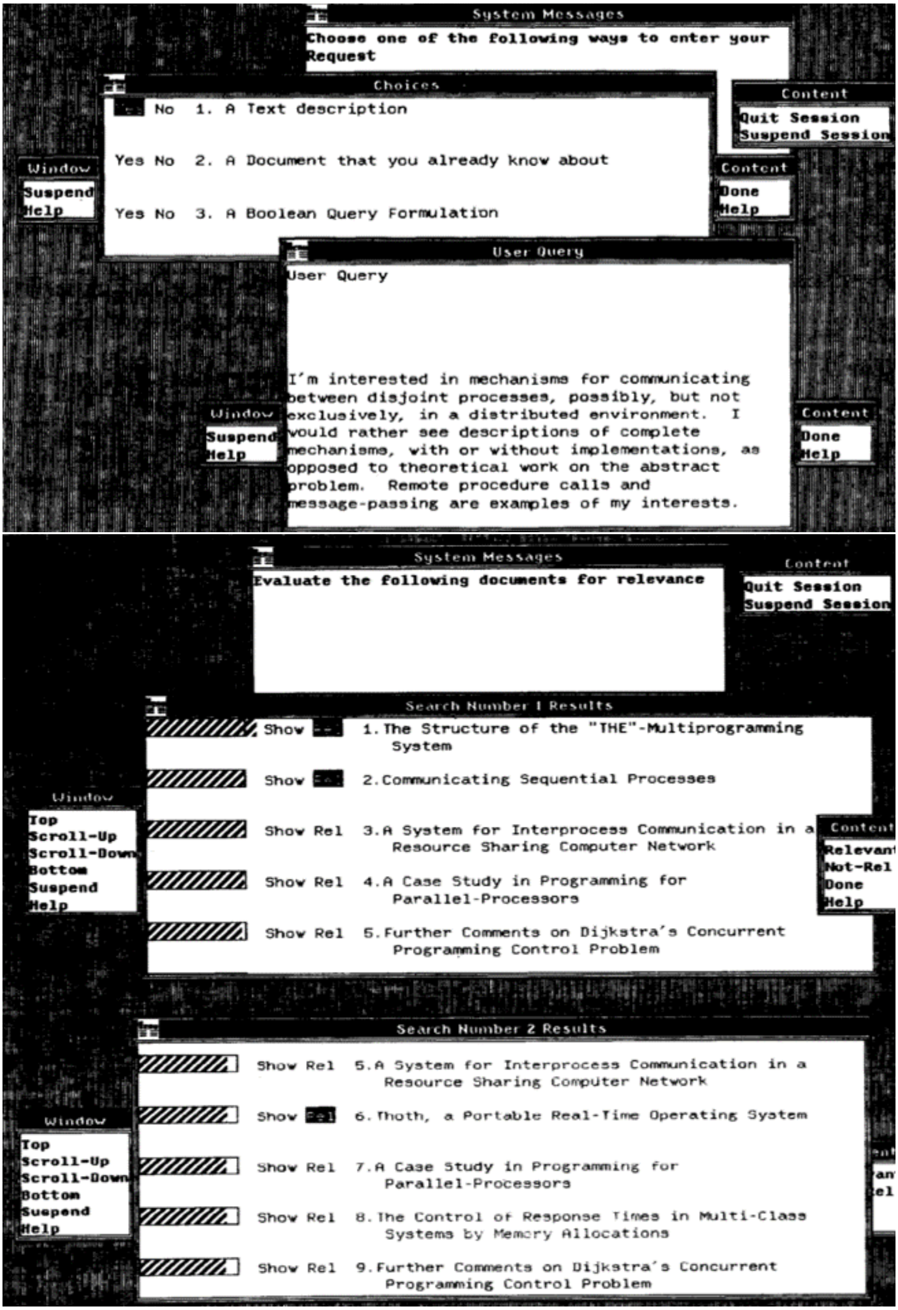

Figure A.2: Snapshots from the I^3R system (Croft and Thompson, 1987).

search engines are mostly based on algorithms designed for retrieving documents for a single query. A main reason for this is due to the complexity of IIR modeling and evaluation. Recently, there has been some promising progress in formal modeling of IIR problems, including the probability ranking principle for IIR (Fuhr, 2008), the economics models for IIR (Azzopardi, 2011), the game theoretic framework for IR (Zhai, 2016), and the interface card model (Zhang and Zhai, 2015). Conversational information seeking is an emerging application of IIR systems and many of the developed IIR models and evaluation methodologies can be extended to CIS systems too. For further reading on approaches for designing and evaluating IIR systems, we refer the reader to the comprehensive survey by Kelly (2009) and the tutorial by Zhai (2020).

A.3 Session-based Information Retrieval

One can put information retrieval tasks in context based on the user's short-term history (Bennett *et al.*, 2012), their long-term history (Keenoy and Levene, 2003), or their situation (Zamani *et al.*, 2017).

> Short-term history is often formulated by the user interactions with the search engine in a short period of time (*e.g.*, a few minutes), referred to as a *search session*. Sessions are different from conversations in that one can pick up and continue a past conversation, while this is not possible in sessions.

Interactions in a session include past queries, retrieved documents, and clicked documents. Therefore, a session can be considered as a period consisting of all interactions for the same information need (Shen *et al.*, 2005). However, this is a strong assumption. In reality, sessions are complex and they are not all alike. Some sessions contain various interactions and query reformulations for a single information need, while some other sessions may involve a series of related simple tasks. Therefore, sessions should be treated differently. This makes modeling search sessions challenging. Existing methods oftentimes relax the assumptions. For instance, Shen *et al.* (2005) assumed that all queries in a session represent the same information need and proposed a model based on the language modeling framework (Ponte and Croft, 1998) for session search tasks. In more detail, they provide a more accurate query language model by interpolating the distribution estimated from the current query, with the ones estimated from the past queries and clicked documents. Bennett *et al.* (2012) introduced a learning to rank approach for session search and defined a number of features that can be used for improving the session search performance in web search. TREC Session Track (Carterette *et al.*, 2016) focused on the development of query formulation during a search session and improving retrieval performance by incorporating knowledge of the session context. Session information can also be used for a number of other information retrieval tasks, such as query suggestion (Sordoni *et al.*, 2015; Dehghani *et al.*, 2017) and clarification (Zamani *et al.*, 2020a).

Whole session evaluation of IR systems is also challenging. Järvelin *et al.* (2008) proposed sDCG, an extension of the nDCG (Järvelin and Kekäläinen, 2002) metric to session search tasks. sDCG basically sums up the nDCG values of all the queries in the session and gives higher weight to the earlier queries. Kanoulas *et al.* (2011) later introduced a normalized variation of sDCG, called nsDCG. Jiang and Allan (2016) conducted a user study to measure the correlation between these metrics and user's opinion. They found that nsDCG has a significant yet weak correlation with the user metrics. They also showed that user's opinions are highly correlated with the performance of the worst and the last queries in the session. More recently, Lipani *et al.* (2019) proposed a user model for session search in which users at each step make a decision to assess the documents in the result list or submit a new query. This user model led to the development of the sRBP metric.

It is clear that session search provides a logical foundation for conversational search tasks, however, there are some fundamental differences that necessitates developing novel models and evaluation methodologies for the conversational search tasks. For instance, since most conversational systems are using limited-bandwidth interfaces, the underlying user models of the aforementioned metrics cannot be extended to conversational search. From the modeling perspective, the type of queries in conversational systems are closer to natural language compared to the session search tasks. In addition, unlike in session search, co-reference and ellipsis resolutions play a central role in conversational search. That being said, we believe that the rich history of IR research on session search would be sometimes quite useful in developing and evaluating conversational search systems.

A.4 Exploratory Search

A significant research effort in interactive IR has focused on *exploratory search* tasks. Exploratory search is an information retrieval task in which the user is unfamiliar with the search task, unsure about the goal, or even unsure about how to complete the task. Users engage in exploratory search with the aim of learning about and exploring a topic – as opposed to known-item/look-up tasks in which users are focused on finding a fact or answering a specific question. Exploratory search refers to a broad set of real-world search tasks that involve learning, investigation, planning, discovery, aggregation, and synthesis (Marchionini, 2006). Exploratory search tasks can be generally categorized as (1) exploratory browsing and (2) focused searching (White and Roth, 2009). Previous work on exploratory search has examined interface features to support users with query refinement and filtering (*e.g.*, faceted search) (Hearst, 2006); tools to help gather and synthesize information (Morris *et al.*, 2008; Donato *et al.*, 2010; Hearst and Degler, 2013); and tools to support collaboration (Golovchinsky *et al.*, 2009).

Natural language conversation is a convenient way for exploratory search tasks. In many exploratory search tasks, users experience difficulties describing their information needs using accurate keyword queries. This is mainly due to a misconception of the topics and/or the document collection. Information seeking conversations would be the natural solution for this problem as natural language conversation is perhaps the most convenient way of human communication and users can express their exploratory search needs quite easily.

Interestingly, many conversations in the TREC CAsT Tracks (Dalton *et al.*, 2019; Dalton *et al.*, 2020a) are basically addressing exploratory information seeking tasks through natural language conversation.

A.5 Dialogue Systems

CIS is also related to dialogue systems. Many concepts used in developing CIS systems were also explored in the context of dialogue systems. Dialogue systems, or conversational agents, refer to computer systems that are intended to converse with humans through natural language. That being said, dialogue systems are not limited to natural language interactions and can benefit from one or more of text, speech, graphics, haptics, gestures, and other modalities. Dialogue systems are mainly categorized as either chatbots (a.k.a. chit-chat dialogue systems) or task-oriented dialogue systems. The former is designed to mimic human conversations mostly for entertainment, while the latter is developed to help the user accomplish a task, *e.g.*, hotel reservation. Task-oriented dialogues are closer to CIS yet with fundamental differences.

Designing and developing dialogue systems require a deep understanding of human conversation. Therefore, the dialogue community spent considerable efforts on extracting and modeling conversations. Jurafsky and Martin (2021) reviewed these properties in detail. For instance, *turn* is perhaps the simplest property – a dialogue is a sequence of turns, each a single contribution from one speaker. *Dialogue acts* is another important property – each dialogue utterance is a type of action performed by the speaker. Different modules in real-world dialogue systems are designed because of this property, such as dialogue act classification. *Grounding* is yet another property of dialogues – acknowledging that dialogue participants understand each other. *Initiative* is the final property we review here. As mentioned in Section 6, it is common in human conversations for initiative to shift back and forth between the participants. For example, in response to a question, a participant can ask for a clarification instead of immediately answering the question. Such interactions are called mixed-initiative. For learning more about dialogue properties and detailed explanations, refer to Jurafsky and Martin (2021, Chapter 24) and McTear *et al.* (2016, Chapter 3).

Dialogue systems have been studied for decades. ELIZA is an early chatbot developed by Weizenbaum (1966) in the 1960s. It is a rule-based dialogue system designed to simulate a Rogerian psychologist. It involves drawing the patient out by reflecting patient's statements back at them. It selects the best match rule for every utterance (regular expression matching) and uses it for producing the next utterance. PARRY is an updated version of ELIZA developed by Colby *et al.* (1971) with a clinical psychology focus, used to study schizophrenia. Besides regular expressions, PARRY models fear and anger and uses these variables to generate utterances. It was the first known system to pass the Turing test, meaning that psychologists could not distinguish its outputs from transcripts of interviews with real paranoids (Colby *et al.*, 1972).

Another successful implementation of dialogue systems in early years was done by the SHRDLU system (Winograd, 1972). SHRDLU provides a natural language interface to a virtual space filled with different blocks. Therefore, SHRDLU users could select and move objects in the virtual space. Given the few number of object types, the action space and vocabulary in SHRDLU conversations are highly limited. The AT&T How May I Help You? (HMIHY) system (Gorin *et al.*, 1997) is also a notable example of dialogue systems developed in the 1990s. HMIHY involved speech recognition, named entity extraction, and intent classification with the goal of call routing. It used a wizard-of-oz approach for data collection and training. It also implemented an active learning algorithm for language understanding.

Dialogue research was later accelerated by the DARPA Communicator Program. For instance, Xu and Rudnicky (2000) developed a language modeling framework for dialogue systems during the Communicator Program. It was designed to support the creation of speech-enabled interfaces that scale across modalities, from speech-only to interfaces that include graphics, maps, pointing and gesture. Recent chatbot systems often use large-scale language models, such as GPT-3 (Brown *et al.*, 2020), in addition to corpus-based approaches that retrieve information from an external corpus in order to produce more sensible utterances.

For task-oriented dialogue systems, Bobrow *et al.* (1977) introduced the GUS architecture in the 1970s. GUS is a frame-based architecture for dialogue systems, where a frame is a kind of knowledge structure representing the information and intention that the system can extract from the user utterances. Thus, frames consist of many slots and dialogue systems need to extract and generate the values of these slots based on the conversation. Architectures similar to or inspired by GUS are still used in real dialogue systems. An alternative to such a modular architecture is end-to-end dialogue systems that do not explicitly model slots and are based on text generation models. We refer the reader to Gao *et al.* (2019, Chapter 4) for recent advances on task-oriented dialogue systems using neural models.

Evaluating dialogue systems is a challenging and widely explored topic. N-grams matching metrics, such as BLEU (Papineni *et al.*, 2002) and ROUGE (Lin, 2004), have been used for dialogue system evaluation. Semantic similarity-based metrics, such as BERT-Score (Zhang *et al.*, 2020b), have also been used. However, research shows that these metrics have several shortcomings (Liu *et al.*, 2016). Using human annotators to evaluate the output of the system and/or using implicit or explicit feedback provided by real users are perhaps the most reliable forms of evaluation for dialogue systems. The PARADISE framework (Walker *et al.*, 1997) for measure overall system success. Developing and evaluating dialogue systems are still active areas of research, we refer the reader to Finch and Choi (2020) for recent work.

A.6 Summary

In this appendix, we briefly reviewed decades of research related to systems and formal models for interactive information retrieval systems, exploratory search, and dialogue systems. Even though the natural language nature of interaction in CIS makes it more complex and many simplifying assumptions made by prior work on IIR cannot be overlooked in the context of CIS systems, many of the concepts that have been developed for IIR can be directly applied to or extended to CIS tasks. The same argument holds for past research on dialogue systems that has been briefly reviewed in the last subsection. Therefore, instead of re-inventing the wheel for various problems in CIS systems, we urge the reader to have a thorough review of the rich literature on IIR and dialogue research, some of which are pointed out in this appendix.

B

A List of Notable CIS Datasets

As discussed in Section 7, data construction for conversational information seeking tasks is challenging yet crucial for advancing the state of the art. Many resources have been developed by the research community in this area. In the following, we provide a non-exhaustive list of notable text-focused resources related to CIS research, in no particular order:

Dataset: TREC CAsT 2019 - 2022 (Dalton *et al.*, 2019)
Task: open domain conversational passage retrieval
Construction: questions written by organizers & passage pooling
Scale: 100+ conversations

Dataset: CoQA (Reddy *et al.*, 2019)
Task: conversational question answering in seven domains
Construction: wizard-of-oz
Scale: 1K+ conversations

Dataset: QuAC (Choi *et al.*, 2018)
Task: conversational question answering about people
Construction: wizard-of-oz
Scale: 10K+ conversations

Dataset: QReCC (Anantha *et al.*, 2021) Task: open domain conversational question answering Construction: crowdsourcing (professional annotators) Scale: 10K+ conversations
Dataset: TopiOCQA (Adlakha *et al.*, 2021) Task: open domain conversational question answering Construction: wizard-of-oz Scale: 1K+ conversations
Dataset: MISC (Thomas *et al.*, 2017) Task: open domain conversational information seeking Construction: spoken human conversations Scale: 10+ conversations
Dataset: Qulac (Aliannejadi *et al.*, 2019) Task: open domain CIS clarification Construction: crowdsourcing Scale: 10K+ clarifications
Datasets: MIMICS (Zamani *et al.*, 2020b) & MIMICS-Duo (Tavakoli *et al.*, 2022) Task: open domain CIS clarification Construction: search logs & crowdsourcing (professional annotators) Scale: 100K+ clarifications
Dataset: RiDial (Li *et al.*, 2018) Task: conversational movie recommendation Construction: wizard-of-oz Scale: 10K+ conversations
Dataset: OpenDialKG (Moon *et al.*, 2019) Tasks: (1) conversational movie and book recommendation, and (2) chit-chat about sports and music Construction: wizard-of-oz Scale: 10k+ conversations

Dataset: SCSdata (Trippas *et al.*, 2017)
Task: spoken conversational search
Construction: wizard-of-oz (lab study)
Scale: 10+ conversations

Dataset: CCPE-M (Radlinski *et al.*, 2019)
Task: conversational movie preference elicitation
Construction: wizard-of-oz
Scale: 100+ conversations

Dataset: Frames (El Asri *et al.*, 2017)
Task: task-oriented dialogue for travel assistance
Construction: wizard-of-oz
Scale: 1000+ conversations

Dataset: KVRET (Eric *et al.*, 2017)
Task: task-oriented dialogue for car driver assistance
Construction: wizard-of-oz
Scale: 1000+ conversations

Dataset: MultiWOZ (Budzianowski *et al.*, 2018)
Task: task-oriented dialogue for multiple domains
Construction: wizard-of-oz
Scale: 1K+ conversations

Dataset: TaskMaster (Byrne *et al.*, 2019)
Task: task-oriented dialogue for multiple domains
Construction: wizard-of-oz and self-dialogue
Scale: 10K+ conversations

Dataset: Wizard of Tasks (Choi *et al.*, 2022)
Task: conversational task assistant for two domains: cooking and home improvement
Construction: wizard-of-oz
Scale: 100+ conversations

Dataset: Wizard of Wikipedia (WoW) (Dinan *et al.*, 2019b)
Task: conversational information seeking about Wikipedia topics
Construction: wizard-of-oz
Scale: 10K+ conversations

Dataset: MSDialog (Qu *et al.*, 2018)
Task: intent detection in conversational information seeking
Construction: crawling Microsoft Community Forum and crowd-sourcing annotations
Scale: 1000+ conversations

Dataset: Ubuntu Dialog Corpus (UDC) (Lowe *et al.*, 2015)
Task: technical conversational information seeking about Ubuntu
Construction: Ubuntu chat logs (human-human conversations)
Scale: 100K+ conversations

References

Abdollahpouri, H., G. Adomavicius, R. Burke, I. Guy, D. Jannach, T. Kamishima, J. Krasnodebski, and L. Pizzato. (2020). "Multi-stakeholder recommendation: Survey and research directions". *User Modeling and User-Adapted Interaction*. 30(1): 127–158.

Adadi, A. and M. Berrada. (2018). "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)". *IEEE access*. 6: 52138–52160.

Adiwardana, D. D. F., M.-T. Luong, D. R. So, J. Hall, N. Fiedel, R. Thoppilan, Z. Yang, A. Kulshreshtha, G. Nemade, Y. Lu, and Q. V. Le. (2020). "Towards a Human-like Open-Domain Chatbot". *ArXiv*. abs/2001.09977.

Adlakha, V., S. Dhuliawala, K. Suleman, H. de Vries, and S. Reddy. (2021). "TopiOCQA: Open-domain Conversational Question Answering with Topic Switching". *Transactions of the Association for Computational Linguistics*. 10: 468–483.

Aliannejadi, M., L. Azzopardi, H. Zamani, E. Kanoulas, P. Thomas, and N. Craswell. (2021a). "Analysing Mixed Initiatives and Search Strategies during Conversational Search". In: *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 16–26. URL: https://doi.org/10.1145/3459637.3482231.

Aliannejadi, M., M. Chakraborty, E. A. Píssola, and F. Crestani. (2020). "Harnessing Evolution of Multi-Turn Conversations for Effective Answer Retrieval". In: *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. 33–42. DOI: 10.1145/3343413.3377968.

Aliannejadi, M., H. Zamani, F. Crestani, and W. B. Croft. (2018a). "In Situ and Context-Aware Target Apps Selection for Unified Mobile Search". In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 1383–1392. DOI: 10.1145/3269206.3271679.

Aliannejadi, M., H. Zamani, F. Crestani, and W. B. Croft. (2018b). "Target Apps Selection: Towards a Unified Search Framework for Mobile Devices". In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 215–224. DOI: 10.1145/3209978.3210039.

Aliannejadi, M., H. Zamani, F. Crestani, and W. B. Croft. (2019). "Asking Clarifying Questions in Open-Domain Information-Seeking Conversations". In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 475–484. DOI: 10.1145/3331184.3331265.

Aliannejadi, M., H. Zamani, F. Crestani, and W. B. Croft. (2021b). "Context-Aware Target Apps Selection and Recommendation for Enhancing Personal Mobile Assistants". *ACM Trans. Inf. Syst.* 39(3). DOI: 10.1145/3447678.

Allen, J. E., C. I. Guinn, and E. Horvitz. (1999). "Mixed-Initiative Interaction". *IEEE Intelligent Systems and their Applications*. 14(5): 14–23.

Anand, A., L. Cavedon, H. Joho, M. Sanderson, and B. Stein. (2020). "Conversational Search (Dagstuhl Seminar 19461)". In: Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Anantha, R., S. Vakulenko, Z. Tu, S. Longpre, S. Pulman, and S. Chappidi. (2021). "Open-Domain Question Answering Goes Conversational via Question Rewriting". In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 520–534. DOI: 10.18653/v1/2021.naacl-main.44.

Andolina, S., V. Orso, H. Schneider, K. Klouche, T. Ruotsalo, L. Gamberini, and G. Jacucci. (2018). "Investigating Proactive Search Support in Conversations". In: *Proceedings of the 2018 Designing Interactive Systems Conference*. 1295–1307. DOI: 10.1145/3196709.3196734.

Anelli, V. W., P. Basile, G. de Melo, F. M. Donini, A. Ferrara, C. Musto, F. Narducci, A. Ragone, and M. Zanker. (2022). "Fourth Knowledge-Aware and Conversational Recommender Systems Workshop (KaRS)". In: *Proceedings of the 16th ACM Conference on Recommender Systems*. 663–666. URL: https://doi.org/10.1145/3523227.3547412.

Arnold, A., G. Dupont, C. Kobus, F. Lancelot, and Y.-H. Liu. (2020). "Perceived Usefulness of Conversational Agents Predicts Search Performance in Aerospace Domain". In: *Proceedings of the 2nd Conference on Conversational User Interfaces*. DOI: 10.1145/3405755.3406172.

Arons, B. (1997). "SpeechSkimmer: a system for interactively skimming recorded speech". *ACM Transactions on Computer-Human Interaction (TOCHI)*. 4(1): 3–38.

Avula, S. (2020). "Characterizing and Understanding User Perception of System Initiative for Conversational Systems to Support Collaborative Search". *PhD thesis*. The University of North Carolina at Chapel Hill.

Avula, S. and J. Arguello. (2020). "Wizard of Oz Interface to Study System Initiative for Conversational Search". In: *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. 447–451. DOI: 10.1145/3343413.3377941.

Avula, S., G. Chadwick, J. Arguello, and R. Capra. (2018). "SearchBots: User Engagement with ChatBots during Collaborative Search". In: *Proceedings of the 2018 Conference on Human Information Interaction and Retrieval*. 52–61. DOI: 10.1145/3176349.3176380.

Azzopardi, L., M. Dubiel, M. Halvey, and J. Dalton. (2018). "Conceptualizing agent-human interactions during the conversational search process". In: *CAIR '18*.

Azzopardi, L. (2011). "The Economics in Interactive Information Retrieval". In: *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 15–24. DOI: 10.1145/2009916.2009923.

Azzopardi, L. (2021). "Cognitive Biases in Search: A Review and Reflection of Cognitive Biases in Information Retrieval". In: *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. 27–37. DOI: 10.1145/3406522.3446023.

Azzopardi, L., M. Aliannejadi, and E. Kanoulas. (2022). "Towards Building Economic Models of Conversational Search". In: *Advances in Information Retrieval*. 31–38.

Bagga, S., A. Piper, and D. Ruths. (2021). ""Are you kidding me?": Detecting Unpalatable Questions on Reddit". In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. 2083–2099.

Baheti, A., A. Ritter, and K. Small. (2020). "Fluent Response Generation for Conversational Question Answering". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 191–207. DOI: 10.18653/v1/2020.acl-main.19.

Baker, C. F., C. J. Fillmore, and J. B. Lowe. (1998). "The Berkeley FrameNet Project". In: *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*. 86–90. DOI: 10.3115/980845.980860.

Balasuriya, S. S., L. Sitbon, A. A. Bayor, M. Hoogstrate, and M. Brereton. (2018). "Use of Voice Activated Interfaces by People with Intellectual Disability". In: *Proceedings of the 30th Australian Conference on Computer-Human Interaction*. 102–112. DOI: 10.1145/3292147.3292161.

Baldauf, M., R. Bösch, C. Frei, F. Hautle, and M. Jenny. (2018). "Exploring Requirements and Opportunities of Conversational User Interfaces for the Cognitively Impaired". In: *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. 119–126. DOI: 10.1145/3236112.3236128.

Balog, K., D. Maxwell, P. Thomas, and S. Zhang. (2022). "Report on the 1st Simulation for Information Retrieval Workshop (Sim4IR 2021) at SIGIR 2021". *SIGIR Forum*. 55(2). URL: https://doi.org/10.1145/3527546.3527559.

Balog, K. and F. Radlinski. (2020). "Measuring Recommendation Explanation Quality: The Conflicting Goals of Explanations". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 329–338. DOI: 10.1145/3397271.3401032.

Banerjee, S. and A. Lavie. (2005). "METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments". In: *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*. 65–72. URL: https://www.aclweb.org/anthology/W05-0909.

Bang, Y., S. Cahyawijaya, N. Lee, W. Dai, D. Su, B. Wilie, H. Lovenia, Z. Ji, T. Yu, W. Chung, Q. V. Do, Y. Xu, and P. Fung. (2023). "A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity". *arXiv preprint arXiv:2302.04023*.

Belkin, N. J. (1980). "Anomalous states of knowledge as a basis for information retrieval". *Canadian Journal of Information Science*. 5: 133–143.

Belkin, N. J., C. Cool, A. Stein, and U. Thiel. (1995). "Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems". *Expert Systems with Applications*. 9(3): 379–395.

Belkin, N. J. and B. H. Kwaundefinednik. (1986). "Using Structural Representation of Anomalous States of Knowledge for Choosing Document Retrieval Strategies". In: *Proceedings of the 9th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 11–22. DOI: 10.1145/253168.253175.

Belkin, N. J. and W. B. Croft. (1992). "Information Filtering and Information Retrieval: Two Sides of the Same Coin?" *Commun. ACM*. 35(12): 29–38. DOI: 10.1145/138859.138861.

Beltagy, I., M. E. Peters, and A. Cohan. (2020). "Longformer: The Long-Document Transformer". *arXiv:2004.05150*.

Bendersky, M. and W. B. Croft. (2008). "Discovering Key Concepts in Verbose Queries". In: *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 491–498. DOI: 10.1145/1390334.1390419.

Bennett, P. N., R. W. White, W. Chu, S. T. Dumais, P. Bailey, F. Borisyuk, and X. Cui. (2012). "Modeling the Impact of Short- and Long-term Behavior on Search Personalization". In: *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 185–194. DOI: 10.1145/2348283.2348312.

Benzeghiba, M., R. De Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fissore, P. Laface, A. Mertins, C. Ris, R. Rose, V. Tyagi, and C. Wellekens. (2007). "Automatic Speech Recognition and Speech Variability: A Review". *Speech Commun.* 49(10–11): 763–786. DOI: 10.1016/j.specom.2007.02.006.

Beutel, A., J. Chen, T. Doshi, H. Qian, L. Wei, Y. Wu, L. Heldt, Z. Zhao, L. Hong, E. H. Chi, and C. Goodrow. (2019). "Fairness in Recommendation Ranking through Pairwise Comparisons". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2212–2220. DOI: 10.1145/3292500.3330745.

Bi, K., Q. Ai, and W. B. Croft. (2021). "Asking Clarifying Questions Based on Negative Feedback in Conversational Search". In: *Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval*. 157–166. URL: https://doi.org/10.1145/3471158.3472232.

Bickmore, T. W., D. Utami, R. Matsuyama, and M. K. Paasche-Orlow. (2016). "Improving access to online health information with conversational agents: a randomized controlled experiment". *Journal of medical Internet research*. 18(1): e5239.

Blanco, H. and F. Ricci. (2013). "Acquiring User Profiles from Implicit Feedback in a Conversational Recommender System". In: *Proceedings of the 7th ACM Conference on Recommender Systems*. 307–310. DOI: 10.1145/2507157.2507217.

Bobrow, D. G., R. M. Kaplan, M. Kay, D. A. Norman, H. Thompson, and T. Winograd. (1977). "GUS, a Frame-Driven Dialog System". *Artif. Intell.* 8(2): 155–173. DOI: 10.1016/0004-3702(77)90018-2.

Boye, J., B. A. Hockey, and M. Rayner. (2000). "Asynchronous dialogue management: Two case-studies". In: *Gotalog: Fourth Workshop on the Semantics and Pragmatics of Dialogue.*

Braslavski, P., D. Savenkov, E. Agichtein, and A. Dubatovka. (2017). "What Do You Mean Exactly? Analyzing Clarification Questions in CQA". In: *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval.* 345–348. DOI: 10.1145/3020165.3022149.

Brennan, S. E. (2012). "Conversation and dialogue". In: *Encyclopedia of the Mind.* Ed. by H. Pashler. SAGE Publications. 202–205.

Brooks, H. M. and N. J. Belkin. (1983). "Using Discourse Analysis for the Design of Information Retrieval Interaction Mechanisms". *SIGIR Forum.* 17(4): 31–47. DOI: 10.1145/1013230.511800.

Brown, T., B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. (2020). "Language Models are Few-Shot Learners". In: *Advances in Neural Information Processing Systems.* Ed. by H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin. Vol. 33. 1877–1901. URL: https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf.

Budzianowski, P. and I. Vulić. (2019). "Hello, It's GPT-2 - How Can I Help You? Towards the Use of Pretrained Language Models for Task-Oriented Dialogue Systems". In: *Proceedings of the 3rd Workshop on Neural Generation and Translation.* 15–22. DOI: 10.18653/v1/D19-5602.

Budzianowski, P., T.-H. Wen, B.-H. Tseng, I. Casanueva, S. Ultes, O. Ramadan, and M. Gasic. (2018). "MultiWOZ-A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing.* 5016–5026.

Bunt, H., V. Petukhova, D. Traum, and J. Alexandersson. (2017). "Dialogue Act Annotation with the ISO 24617-2 Standard". In: *Multimodal Interaction with W3C Standards: Toward Natural User Interfaces to Everything*. Ed. by D. A. Dahl. 109–135. DOI: 10.1007/978-3-319-42816-1_6.

Büschel, W., A. Mitschick, and R. Dachselt. (2018). "Here and Now: Reality-Based Information Retrieval: Perspective Paper". In: *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. 171–180. DOI: 10.1145/3176349.3176384.

Buscher, G., J. Gwizdka, J. Teevan, N. J. Belkin, R. Bierig, L. van Elst, and J. Jose. (2009). "SIGIR 2009 Workshop on Understanding the User: Logging and Interpreting User Interactions in Information Search and Retrieval". *SIGIR Forum*. 43(2): 57–62. DOI: 10.1145/1670564.1670574.

Byrne, B., K. Krishnamoorthi, C. Sankar, A. Neelakantan, B. Goodrich, D. Duckworth, S. Yavuz, A. Dubey, K.-Y. Kim, and A. Cedilnik. (2019). "Taskmaster-1: Toward a Realistic and Diverse Dialog Dataset". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 4506–4517.

Car, L. T., D. A. Dhinagaran, B. M. Kyaw, T. Kowatsch, S. Joty, Y.-L. Theng, and R. Atun. (2020). "Conversational agents in health care: scoping review and conceptual analysis". *Journal of medical Internet research*. 22(8): e17158.

Carterette, B., P. Clough, M. Hall, E. Kanoulas, and M. Sanderson. (2016). "Evaluating Retrieval over Sessions: The TREC Session Track 2011-2014". In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 685–688. DOI: 10.1145/2911451.2914675.

Chen, D., A. Fisch, J. Weston, and A. Bordes. (2017). "Reading Wikipedia to Answer Open-Domain Questions". In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1870–1879. DOI: 10.18653/v1/P17-1171.

Chen, L., G. Chen, and F. Wang. (2015). “Recommender Systems Based on User Reviews: The State of the Art”. *User Modeling and User-Adapted Interaction*. 25(2): 99–154. DOI: 10.1007/s11257-015-9155-5.

Chen, S., Z. Chen, X. Deng, A. Lewis, L. Mo, S. Stevens, Z. Wang, X. Yue, T. Zhang, Y. Su, and H. Sun. (2022). “Bootstrapping a user-centered task-oriented dialogue system”. In: *Alexa Prize TaskBot Challenge Proceedings*.

Chen, Y., L. Wu, and M. J. Zaki. (2021a). “GraphFlow: Exploiting Conversation Flow with Graph Neural Networks for Conversational Machine Comprehension”. In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*.

Chen, Z., X. Wang, X. Xie, M. Parsana, A. Soni, X. Ao, and E. Chen. (2021b). “Towards Explainable Conversational Recommendation”. In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*.

Chiang, T.-R., H.-T. Ye, and Y.-N. Chen. (2020). “An Empirical Study of Content Understanding in Conversational Question Answering”. *ArXiv*. abs/1909.10743.

Choi, B.-J., J. Hong, D. Park, and S. W. Lee. (2020). “F^2-Softmax: Diversifying Neural Text Generation via Frequency Factorized Softmax”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. 9167–9182. DOI: 10.18653/v1/2020.emnlp-main.737.

Choi, E., H. He, M. Iyyer, M. Yatskar, W.-t. Yih, Y. Choi, P. Liang, and L. Zettlemoyer. (2018). “QuAC: Question Answering in Context”. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2174–2184.

Choi, J. I., S. Kuzi, N. Vedula, J. Zhao, G. Castellucci, M. Collins, S. Malmasi, O. Rokhlenko, and E. Agichtein. (2022). “Wizard of Tasks: A Novel Conversational Dataset for Solving Real-World Tasks in Conversational Settings”. In: *Proceedings of the 29th International Conference on Computational Linguistics*. 3514–3529. URL: https://aclanthology.org/2022.coling-1.310.

Christakopoulou, K., F. Radlinski, and K. Hofmann. (2016). "Towards Conversational Recommender Systems". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 815–824. DOI: 10.1145/2939672.2939746.

Christmann, P., R. Saha Roy, A. Abujabal, J. Singh, and G. Weikum. (2019). "Look before You Hop: Conversational Question Answering over Knowledge Graphs Using Judicious Context Expansion". In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 729–738. DOI: 10.1145/3357384.3358016.

Christmann, P., R. Saha Roy, and G. Weikum. (2022). "Conversational Question Answering on Heterogeneous Sources". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 144–154. DOI: 10.1145/3477495.3531815.

Chuklin, A., A. Severyn, J. R. Trippas, E. Alfonseca, H. Silen, and D. Spina. (2018). "Prosody Modifications for Question-Answering in Voice-Only Settings". *arXiv preprint arXiv:1806.03957*: 1–5.

Chuklin, A., A. Severyn, J. R. Trippas, E. Alfonseca, H. Silen, and D. Spina. (2019). "Using Audio Transformations to Improve Comprehension in Voice Question Answering". In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. Ed. by F. Crestani, M. Braschler, J. Savoy, A. Rauber, H. Müller, D. E. Losada, G. Heinatz Bürki, L. Cappellato, and N. Ferro. 164–170.

Church, K. and N. Oliver. (2011). "Understanding Mobile Web and Mobile Search Use in Today's Dynamic Mobile Landscape". In: *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*. 67–76. DOI: 10.1145/2037373.2037385.

Clark, D. (1988). "The Design Philosophy of the DARPA Internet Protocols". *SIGCOMM Comput. Commun. Rev.* 18(4): 106–114. DOI: 10.1145/52325.52336.

Clarke, C. (2019). "WaterlooClarke at the TREC 2019 Conversational Assistant Track". In: *TREC*.

Clarke, C. L. A., E. Agichtein, S. Dumais, and R. W. White. (2007). "The Influence of Caption Features on Clickthrough Patterns in Web Search". In: *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 135–142. DOI: 10.1145/1277741.1277767.

Cleverdon, C. and M. Kean. (1968). "Factors Determining the Performance of Indexing Systems". Aslib Cranfield Research Project, Cranfield, England.

Coden, A., D. Gruhl, N. Lewis, and P. N. Mendes. (2015). "Did you mean A or B? Supporting Clarification Dialog for Entity Disambiguation". In: *Joint Proceedings of the 1st International Workshop on Summarizing and Presenting Entities and Ontologies and the 3rd International Workshop on Human Semantic Web Interfaces (SumPre 2015, HSWI 2015) co-located with the 12th Extended Semantic Web Conference (ESWC 2015)*. Vol. 1556.

Colby, K. M., F. D. Hilf, S. Weber, and H. C. Kraemer. (1972). "Turing-like Indistinguishability Tests for the Validation of a Computer Simulation of Paranoid Processes". *Artif. Intell.* 3(1): 199–221. DOI: 10.1016/0004-3702(72)90049-5.

Colby, K. M., S. Weber, and F. D. Hilf. (1971). "Artificial Paranoia". *Artificial Intelligence*. 2(1): 1–25. DOI: https://doi.org/10.1016/0004-3702(71)90002-6.

Cooper, A. (2008). "A Survey of Query Log Privacy-Enhancing Techniques from a Policy Perspective". *ACM Trans. Web*. 2(4). DOI: 10.1145/1409220.1409222.

Croft, W. B., D. Metzler, and T. Strohman. (2010). *Search engines: Information retrieval in practice*. Vol. 520. Addison-Wesley Reading.

Croft, W. B. and R. H. Thompson. (1987). "I^3R: A New Approach to the Design of Document Retrieval Systems". *J. Am. Soc. Inf. Sci.* 38(6): 389–404.

Cross, S., A. Mourad, G. Zuccon, and B. Koopman. (2021). "Search Engines vs. Symptom Checkers: A Comparison of Their Effectiveness for Online Health Advice". In: *Proceedings of the Web Conference 2021*. 206–216. DOI: 10.1145/3442381.3450140.

Cucerzan, S. (2007). “Large-Scale Named Entity Disambiguation Based on Wikipedia Data”. In: *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*. 708–716. URL: https://aclanthology.org/D07-1074.

Culpepper, J. S., F. Diaz, and M. D. Smucker. (2018). “Research Frontiers in Information Retrieval: Report from the Third Strategic Workshop on Information Retrieval in Lorne (SWIRL 2018)”. *SIGIR Forum*. 52(1): 34–90. DOI: 10.1145/3274784.3274788.

Cutrell, E. and Z. Guan. (2007). “What Are You Looking for? An Eye-Tracking Study of Information Usage in Web Search”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 407–416. DOI: 10.1145/1240624.1240690.

Dalton, J., S. Fischer, P. Owoicho, F. Radlinski, F. Rossetto, J. R. Trippas, and H. Zamani. (2022). “Conversational Information Seeking: Theory and Application”. In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 3455–3458. DOI: 10.1145/3477495.3532678.

Dalton, J., F. Radlinski, F. Rossetto, J. R. Trippas, and H. Zamani. (2023). “Conversational Information Seeking: Theory and Application”. In: *Proceedings of the Web Conference 2023*.

Dalton, J., C. Xiong, and J. Callan. (2019). “TREC CAsT 2019: The Conversational Assistance Track Overview”. In: *TREC*.

Dalton, J., C. Xiong, and J. Callan. (2020a). “TREC CAsT 2020: The Conversational Assistance Track Overview”. In: *TREC*.

Dalton, J., C. Xiong, and J. Callan. (2021). “TREC CAsT 2021: The Conversational Assistance Track Overview”. In: *TREC*.

Dalton, J., C. Xiong, V. Kumar, and J. Callan. (2020b). “CAsT-19: A Dataset for Conversational Information Seeking”. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1985–1988. DOI: 10.1145/3397271.3401206.

Daronnat, S., L. Azzopardi, M. Halvey, and M. Dubiel. (2020). "Impact of Agent Reliability and Predictability on Trust in Real Time Human-Agent Collaboration". In: *Proceedings of the 8th International Conference on Human-Agent Interaction*. 131–139. DOI: 10.1145/3406499.3415063.

Dehghani, M., H. Azarbonyad, J. Kamps, and M. de Rijke. (2019). "Learning to Transform, Combine, and Reason in Open-Domain Question Answering". In: *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. 681–689. DOI: 10.1145/3289600.3291012.

Dehghani, M., S. Rothe, E. Alfonseca, and P. Fleury. (2017). "Learning to Attend, Copy, and Generate for Session-Based Query Suggestion". In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 1747–1756. DOI: 10.1145/3132847.3133010.

Deits, R., S. Tellex, P. Thaker, D. Simeonov, T. Kollar, and N. Roy. (2013). "Clarifying Commands with Information-Theoretic Human-Robot Dialog". *J. Hum.-Robot Interact.* 2(2): 58–79. DOI: 10.5898/JHRI.2.2.Deits.

Del Tredici, M., G. Barlacchi, X. Shen, W. Cheng, and A. de Gispert. (2021). "Question Rewriting for Open-Domain Conversational QA: Best Practices and Limitations". In: *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 2974–2978. DOI: 10.1145/3459637.3482164.

Deldjoo, Y., J. R. Trippas, and H. Zamani. (2021). "Towards Multi-Modal Conversational Information Seeking". In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1577–1587. DOI: 10.1145/3404835.3462806.

Derboven, J., J. Huyghe, and D. De Grooff. (2014). "Designing Voice Interaction for People with Physical and Speech Impairments". In: *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. 217–226. DOI: 10.1145/2639189.2639252.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 4171–4186. DOI: 10.18653/v1/N19-1423.

Di Noia, T., N. Tintarev, P. Fatourou, and M. Schedl. (2022). "Recommender Systems under European AI Regulations". *Commun. ACM*. 65(4): 69–73. DOI: 10.1145/3512728.

Diaz, F. and D. Metzler. (2006). "Improving the Estimation of Relevance Models Using Large External Corpora". In: *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 154–161. DOI: 10.1145/1148170.1148200.

Dinan, E., V. Logacheva, V. Malykh, A. H. Miller, K. Shuster, J. Urbanek, D. Kiela, A. Szlam, I. Serban, R. Lowe, S. Prabhumoye, A. Black, A. I. Rudnicky, J. Williams, J. Pineau, M. Burtsev, and J. Weston. (2019a). "The Second Conversational Intelligence Challenge (ConvAI2)". *ArXiv*. abs/1902.00098.

Dinan, E., S. Roller, K. Shuster, A. Fan, M. Auli, and J. Weston. (2019b). "Wizard of Wikipedia: Knowledge-Powered Conversational Agents". In: *International Conference on Learning Representations*. URL: https://openreview.net/forum?id=r1l73iRqKm.

Donato, D., F. Bonchi, T. Chi, and Y. Maarek. (2010). "Do You Want to Take Notes? Identifying Research Missions in Yahoo! Search Pad". In: *Proceedings of the 19th International Conference on World Wide Web*. 321–330. DOI: 10.1145/1772690.1772724.

Du, J., Z. Zhang, J. Yan, Y. Cui, and Z. Chen. (2010). "Using Search Session Context for Named Entity Recognition in Query". In: *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 765–766. DOI: 10.1145/1835449.1835605.

Dubiel, M., M. Halvey, L. Azzopardi, D. Anderson, and S. Daronnat. (2020a). "Conversational strategies: impact on search performance in a goal-oriented task". In: *The Third International Workshop on Conversational Approaches to Information Retrieval*.

Dubiel, M., M. Halvey, L. Azzopardi, and S. Daronnat. (2018). "Investigating how conversational search agents affect user's behaviour, performance and search experience". In: *The Second International Workshop on Conversational Approaches to Information Retrieval.*

Dubiel, M., M. Halvey, P. O. Gallegos, and S. King. (2020b). "Persuasive Synthetic Speech: Voice Perception and User Behaviour". In: *Proceedings of the 2nd Conference on Conversational User Interfaces.* DOI: 10.1145/3405755.3406120.

Dumais, S., R. Jeffries, D. M. Russell, D. Tang, and J. Teevan. (2014). "Understanding User Behavior Through Log Data and Analysis". In: *Ways of Knowing in HCI.* Ed. by J. S. Olson and W. A. Kellogg. Springer New York. 349–372. DOI: 10.1007/978-1-4939-0378-8_14.

Dušek, O., J. Novikova, and V. Rieser. (2018). "Findings of the E2E NLG Challenge". In: *Proceedings of the 11th International Conference on Natural Language Generation.* 322–328. DOI: 10.18653/v1/W18-6539.

Dwivedi, S. K. and V. Singh. (2013). "Research and reviews in question answering system". *Procedia Technology.* 10: 417–424.

Dziri, N., H. Rashkin, T. Linzen, and D. Reitter. (2022). "Evaluating Attribution in Dialogue Systems: The BEGIN Benchmark". *Transactions of the Association for Computational Linguistics.* 10: 1066–1083. DOI: 10.1162/tacl_a_00506.

Ekstrand, M. D., A. Chaney, P. Castells, R. Burke, D. Rohde, and M. Slokom. (2021). "SimuRec: Workshop on Synthetic Data and Simulation Methods for Recommender Systems Research". In: *Proceedings of the 15th ACM Conference on Recommender Systems.* 803–805. DOI: 10.1145/3460231.3470938.

Ekstrand, M. D., A. Das, R. Burke, and F. Diaz. (2022). "Fairness in Information Access Systems". *Foundations and Trends® in Information Retrieval.* 16(1-2): 1–177. URL: http://dx.doi.org/10.1561/1500000079.

El Asri, L., H. Schulz, S. Sharma, J. Zumer, J. Harris, E. Fine, R. Mehrotra, and K. Suleman. (2017). "Frames: a corpus for adding memory to goal-oriented dialogue systems". In: *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue.* 207–219. DOI: 10.18653/v1/W17-5526.

Elgohary, A., D. Peskov, and J. Boyd-Graber. (2019). "Can You Unpack That? Learning to Rewrite Questions-in-Context". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 5918–5924. DOI: 10.18653/v1/D19-1605.

Erbacher, P., L. Denoyer, and L. Soulier. (2022). "Interactive Query Clarification and Refinement via User Simulation". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2420–2425. DOI: 10.1145/3477495.3531871.

Eric, M., L. Krishnan, F. Charette, and C. D. Manning. (2017). "Key-Value Retrieval Networks for Task-Oriented Dialogue". In: *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*. 37–49. DOI: 10.18653/v1/W17-5506.

Feild, H. A., J. Allan, and R. Jones. (2010). "Predicting Searcher Frustration". In: *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 34–41. DOI: 10.1145/1835449.1835458.

Ferguson, G. and J. F. Allen. (1998). "TRIPS: An Integrated Intelligent Problem-Solving Assistant". In: *Proceedings of the Fifteenth National Conference on Artificial Intelligence and Tenth Innovative Applications of Artificial Intelligence Conference*. 567–572.

Ferreira, R., D. Tavares, D. Silva, F. Vicente, M. Bonito, G. Gonçalves, R. Margarido, P. Figueiredo, H. Rodrigues, D. Semedo, and J. Magalhaes. (2022). "TWIZ: A conversational Task Wizard with multimodal curiosity-exploration". In: *Alexa Prize TaskBot Challenge Proceedings*.

Finch, S. E. and J. D. Choi. (2020). "Towards Unified Dialogue System Evaluation: A Comprehensive Analysis of Current Evaluation Protocols". In: *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 236–245.

Firdaus, M., N. Thakur, and A. Ekbal. (2021). "Aspect-Aware Response Generation for Multimodal Dialogue System". *ACM Trans. Intell. Syst. Technol.* 12(2). DOI: 10.1145/3430752.

Fono, D. and R. Baecker. (2006). "Structuring and Supporting Persistent Chat Conversations". In: *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*. 455–458. DOI: 10.1145/1180875.1180944.

Formal, T., C. Lassance, B. Piwowarski, and S. Clinchant. (2021). "SPLADE v2: Sparse Lexical and Expansion Model for Information Retrieval". *ArXiv*. abs/2109.10086.

Foster, M. E., R. Alami, O. Gestranius, O. Lemon, M. Niemelä, J.-M. Odobez, and A. K. Pandey. (2016). "The MuMMER Project: Engaging Human-Robot Interaction in Real-World Public Spaces". In: *Social Robotics*. Ed. by A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, and H. He. 753–763.

Fraser, N. (1998). "Assessment of interactive systems". In: *Handbook of standards and resources for spoken language systems*. Ed. by D. Gibbon, R. Moore, and R. Winski. 564–615.

Freed, A. R. (2021). *Conversational AI: Chatbots that work*. Manning Publications Co.

Frieder, O., I. Mele, C. I. Muntean, F. M. Nardini, R. Perego, and N. Tonellotto. (2022). "Caching Historical Embeddings in Conversational Search". *ACM Trans. Web*. Dec. DOI: 10.1145/3578519.

Fu, X., E. Yilmaz, and A. Lipani. (2022). "Evaluating the Cranfield Paradigm for Conversational Search Systems". In: *Proceedings of the 2022 ACM SIGIR International Conference on Theory of Information Retrieval*. 275–280. DOI: 10.1145/3539813.3545126.

Fuhr, N. (2008). "A Probability Ranking Principle for Interactive Information Retrieval". *Inf. Retr.* 11(3): 251–265. DOI: 10.1007/s10791-008-9045-0.

Gao, C., W. Lei, X. He, M. de Rijke, and T.-S. Chua. (2021a). "Advances and challenges in conversational recommender systems: A survey". *AI Open*. 2: 100–126. DOI: https://doi.org/10.1016/j.aiopen.2021.06.002.

Gao, J., M. Galley, and L. Li. (2019). "Neural Approaches to Conversational AI". *Foundations and Trends® in Information Retrieval*. 13(2-3): 127–298. DOI: 10.1561/1500000074.

Gao, J., C. Xiong, P. Bennett, and N. Craswell. (2023). *Neural Approaches to Conversational Information Retrieval*. Springer Cham. URL: https://doi.org/10.1007/978-3-031-23080-6.

Gao, Y., J. Li, M. R. Lyu, and I. King. (2021b). “Open-Retrieval Conversational Machine Reading”. *ArXiv*. abs/2102.08633.

Gao, Y., C.-S. Wu, J. Li, S. Joty, S. C. Hoi, C. Xiong, I. King, and M. Lyu. (2020). “Discern: Discourse-Aware Entailment Reasoning Network for Conversational Machine Reading”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2439–2449. DOI: 10.18653/v1/2020.emnlp-main.191.

Ge, Y., S. Liu, R. Gao, Y. Xian, Y. Li, X. Zhao, C. Pei, F. Sun, J. Ge, W. Ou, and Y. Zhang. (2021). “Towards Long-Term Fairness in Recommendation”. In: *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 445–453. DOI: 10.1145/3437963.3441824.

Gekhman, Z., N. Oved, O. Keller, I. Szpektor, and R. Reichart. (2022). “On the Robustness of Dialogue History Representation in Conversational Question Answering: A Comprehensive Study and a New Prompt-based Method”. *ArXiv*. abs/2206.14796.

Gemmell, C. and J. Dalton. (2020). “Glasgow Representation and Information Learning Lab (GRILL) at the Conversational Assistance Track 2020”. In: *TREC*.

Gemmell, C., S. Fischer, I. Mackie, P. Owoicho, F. Rossetto, and J. Dalton. (2022). “GRILLBot: A flexible conversational agent for solving complex real-world tasks”. In: *Alexa Prize TaskBot Challenge Proceedings*.

Gerritse, E. J., F. Hasibi, and A. P. de Vries. (2020). “Bias in Conversational Search: The Double-Edged Sword of the Personalized Knowledge Graph”. In: *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*. 133–136. DOI: 10.1145/3409256.3409834.

Gibbon, D., R. Moore, and R. Winski. (1997). *Handbook of Standards and Resources for Spoken Language Systems*. de Gruyter.

Gibson, W. (2009). "Intercultural communication online: Conversation analysis and the investigation of asynchronous written discourse". In: *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*. Vol. 10. No. 1.

Goes, P., N. Ilk, W. T. Yue, and J. L. Zhao. (2012). "Live-Chat Agent Assignments to Heterogeneous e-Customers under Imperfect Classification". *ACM Trans. Manage. Inf. Syst.* 2(4). DOI: 10.1145/2070710.2070715.

Golovchinsky, G., P. Qvarfordt, and J. Pickens. (2009). "Collaborative Information Seeking". *Computer*. 42(3): 47–51. DOI: 10.1109/MC.2009.73.

Gooda Sahib, N., A. Tombros, and T. Stockman. (2015). "Evaluating a Search Interface for Visually Impaired Searchers". *J. Assoc. Inf. Sci. Technol.* 66(11): 2235–2248.

Gorin, A. L., G. Riccardi, and J. H. Wright. (1997). "How May I Help You?" *Speech Commun.* 23(1–2): 113–127. DOI: 10.1016/S0167-6393(97)00040-X.

Gosper, S., J. R. Trippas, H. Richards, F. Allison, C. Sear, S. Khorasani, and F. Mattioli. (2021). "Understanding the Utility of Digital Flight Assistants: A Preliminary Analysis". In: *CUI 2021 - 3rd Conference on Conversational User Interfaces*. DOI: 10.1145/3469595.3469627.

Gottardi, A., O. Ipek, G. Castellucci, S. Hu, L. Vaz, Y. Lu, A. Khatri, A. Chadha, D. Zhang, S. Sahai, P. Dwivedi, H. Shi, L. Hu, A. Huang, L. Dai, B. Yang, V. Somani, P. Rajan, R. Rezac, M. Johnston, S. Stiff, L. Ball, D. Carmel, Y. Liu, D. Hakkani-Tür, O. Rokhlenko, K. Bland, E. Agichtein, R. Ghanadan, and Y. Maarek. (2022). "Alexa, let's work together: Introducing the first Alexa Prize TaskBot Challenge on conversational task assistance". In: *Alexa Prize TaskBot Challenge Proceedings*. URL: https://www.amazon.science/publications/alexa-lets-work-together-introducing-the-first-alexa-prize-taskbot-challenge-on-conversational-task-assistance.

Green, B. F., A. K. Wolf, C. Chomsky, and K. Laughery. (1961). "Baseball: An Automatic Question-Answerer". In: *Papers Presented at the May 9-11, 1961, Western Joint IRE-AIEE-ACM Computer Conference*. 219–224. DOI: 10.1145/1460690.1460714.

Gunasekara, R. C., S. Kim, L. F. D'Haro, A. Rastogi, Y. Chen, M. Eric, B. Hedayatnia, K. Gopalakrishnan, Y. Liu, C. Huang, D. Hakkani-Tür, J. Li, Q. Zhu, L. Luo, L. Liden, K. Huang, S. Shayandeh, R. Liang, B. Peng, Z. Zhang, S. Shukla, M. Huang, J. Gao, S. Mehri, Y. Feng, C. Gordon, S. H. Alavi, D. R. Traum, M. Eskénazi, A. Beirami, E. Cho, P. A. Crook, A. De, A. Geramifard, S. Kottur, S. Moon, S. Poddar, and R. Subba. (2020). "Overview of the Ninth Dialog System Technology Challenge: DSTC9". *CoRR*. abs/2011.06486. URL: https://arxiv.org/abs/2011.06486.

Guo, D., D. Tang, N. Duan, M. Zhou, and J. Yin. (2018). "Dialog-to-Action: Conversational Question Answering Over a Large-Scale Knowledge Base". In: *Advances in Neural Information Processing Systems*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Vol. 31. URL: https://proceedings.neurips.cc/paper/2018/file/d63fbf8c3173730f82b150c5ef38b8ff-Paper.pdf.

Guo, J., Y. Fan, L. Pang, L. Yang, Q. Ai, H. Zamani, C. Wu, W. B. Croft, and X. Cheng. (2020). "A Deep Look into neural ranking models for information retrieval". *Information Processing & Management*. 57(6): 102067. DOI: https://doi.org/10.1016/j.ipm.2019.102067.

Gupta, A., D. Basu, R. Ghantasala, S. Qiu, and U. Gadiraju. (2022). "To Trust or Not To Trust: How a Conversational Interface Affects Trust in a Decision Support System". In: *Proceedings of the ACM Web Conference 2022*. 3531–3540. DOI: 10.1145/3485447.3512248.

Harabagiu, S. M., D. I. Moldovan, C. Clark, M. Bowden, A. Hickl, and P. Wang. (2005). "Employing Two Question Answering Systems in TREC 2005". In: *TREC*.

Hashemi, H., M. Aliannejadi, H. Zamani, and W. B. Croft. (2019). "ANTIQUE: A Non-factoid Question Answering Benchmark". In: *Proceedings of the 2019 European Conference on Information Retrieval*. 166–173.

Hashemi, H., H. Zamani, and W. B. Croft. (2020). "Guided Transformer: Leveraging Multiple External Sources for Representation Learning in Conversational Search". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1131–1140. DOI: 10.1145/3397271.3401061.

Hashimoto, T. B., K. Guu, Y. Oren, and P. Liang. (2018). "A Retrieve-and-Edit Framework for Predicting Structured Outputs". In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. 10073–10083.

Hassan Awadallah, A., R. Gurunath Kulkarni, U. Ozertem, and R. Jones. (2015). "Characterizing and Predicting Voice Query Reformulation". In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. 543–552. DOI: 10.1145/2806416.2806491.

Hattimare, A., A. Dharawat, Y. Khan, Y.-C. Lien, C. Samarinas, G. Z. Wei, Y. Yang, and H. Zamani. (2022). "Maruna Bot: An extensible retrieval-focused framework for task-oriented dialogues". In: *Alexa Prize TaskBot Challenge Proceedings*.

Hauptmann, A., J. Magalhaes, R. G. Sousa, and J. P. Costeira. (2020). "MuCAI'20: 1st International Workshop on Multimodal Conversational AI". In: *Proceedings of the 28th ACM International Conference on Multimedia*. 4767–4768. DOI: 10.1145/3394171.3421900.

Hawking, D. (2004). "Challenges in Enterprise Search". In: *Proceedings of the 15th Australasian Database Conference - Volume 27*. 15–24.

Hawking, D., B. Billerbeck, P. Thomas, and N. Craswell. (2020). "Simulating Information Retrieval Test Collections". *Synthesis Lectures on Information Concepts, Retrieval, and Services*. 12(2): 1–184. URL: https://doi.org/10.2200/S01043ED1V01Y202008ICR071.

He, J., P. Duboue, and J.-Y. Nie. (2012). "Bridging the gap between intrinsic and perceived relevance in snippet generation". In: *Proceedings of COLING 2012*. 1129–1146.

He, Z., H. Zhao, T. Yu, S. Kim, F. Du, and J. McAuley. (2022). "Bundle MCR: Towards Conversational Bundle Recommendation". In: *Proceedings of the 16th ACM Conference on Recommender Systems*. 288–298. DOI: 10.1145/3523227.3546755.

Hearst, M. A. (2006). "Clustering versus faceted categories for information exploration". *Communications of the ACM*. 49(4): 59–61.

Hearst, M. A. (2009). *Search User Interfaces*. Cambridge University Press.

Hearst, M. A. and D. Degler. (2013). "Sewing the Seams of Sensemaking: A Practical Interface for Tagging and Organizing Saved Search Results". In: *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval*. DOI: 10.1145/2528394.2528398.

Henderson, M., P. Budzianowski, I. Casanueva, S. Coope, D. Gerz, G. Kumar, N. Mrkšić, G. Spithourakis, P.-H. Su, I. Vulić, and T.-H. Wen. (2019). "A Repository of Conversational Datasets". In: *Proceedings of the First Workshop on NLP for Conversational AI*. 1–10. DOI: 10.18653/v1/W19-4101.

Henderson, M., I. Casanueva, N. Mrkšić, P.-H. Su, T.-H. Wen, and I. Vulić. (2020). "ConveRT: Efficient and Accurate Conversational Representations from Transformers". In: *Findings of the Association for Computational Linguistics: EMNLP 2020*. 2161–2174. DOI: 10.18653/v1/2020.findings-emnlp.196.

Hochreiter, S. and J. Schmidhuber. (1997). "Long short-term memory". *Neural computation*. 9(8): 1735–1780.

Horvitz, E. (1999). "Principles of Mixed-initiative User Interfaces". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 159–166. DOI: 10.1145/302979.303030.

Hu, C., S. Huang, Y. Zhang, and Y. Liu. (2022). "Learning to Infer User Implicit Preference in Conversational Recommendation". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 256–266. DOI: 10.1145/3477495.3531844.

Huang, H.-Y., E. Choi, and W.-T. Yih. (2019). "FlowQA: Grasping Flow in History for Conversational Machine Comprehension". In: *International Conference on Learning Representations*. URL: https://openreview.net/forum?id=ByftGnR9KX.

Humeau, S., K. Shuster, M.-A. Lachaux, and J. Weston. (2020). "Poly-encoders: Architectures and Pre-training Strategies for Fast and Accurate Multi-sentence Scoring". In: *International Conference on Learning Representations*. URL: https://openreview.net/forum?id=SkxgnnNFvH.

Ie, E., C.-w. Hsu, M. Mladenov, V. Jain, S. Narvekar, J. Wang, R. Wu, and C. Boutilier. (2019). "RecSim: A Configurable Simulation Platform for Recommender Systems". arXiv: 1909.04847 [cs.LG].

Iovine, A. (2020). "Conversational Agents for Recommender Systems". In: *Fourteenth ACM Conference on Recommender Systems*. 758–763. DOI: 10.1145/3383313.3411453.

Iyyer, M., W.-t. Yih, and M.-W. Chang. (2017). "Search-based neural structured learning for sequential question answering". In: *In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1821–1831.

Izacard, G. and E. Grave. (2021). "Leveraging Passage Retrieval with Generative Models for Open Domain Question Answering". In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. 874–880. DOI: 10.18653/v1/2021.eacl-main.74.

Jannach, D., L. Lerche, and M. Zanker. (2018). "Recommending Based on Implicit Feedback". In: *Social Information Access: Systems and Technologies*. Ed. by P. Brusilovsky and D. He. 510–569. DOI: 10.1007/978-3-319-90092-6_14.

Jannach, D., A. Manzoor, W. Cai, and L. Chen. (2021a). "A Survey on Conversational Recommender Systems". *ACM Comput. Surv.* 54(5). DOI: 10.1145/3453154.

Jannach, D., A. Manzoor, W. Cai, and L. Chen. (2021b). "A Survey on Conversational Recommender Systems". *ACM Comput. Surv.* 54(5). DOI: 10.1145/3453154.

Järvelin, K. and J. Kekäläinen. (2002). "Cumulated Gain-Based Evaluation of IR Techniques". *ACM Trans. Inf. Syst.* 20(4): 422–446. DOI: 10.1145/582415.582418.

Järvelin, K., S. L. Price, L. M. L. Delcambre, and M. L. Nielsen. (2008). "Discounted Cumulated Gain Based Evaluation of Multiple-Query IR Sessions". In: *Proceedings of the IR Research, 30th European Conference on Advances in Information Retrieval*. 4–15.

Jiang, J. and J. Allan. (2016). "Correlation Between System and User Metrics in a Session". In: *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*. 285–288. DOI: 10.1145/2854946.2855005.

Jiang, J., A. Hassan Awadallah, R. Jones, U. Ozertem, I. Zitouni, R. Gurunath Kulkarni, and O. Z. Khan. (2015). "Automatic Online Evaluation of Intelligent Assistants". In: *Proceedings of the 24th International Conference on World Wide Web*. 506–516. DOI: 10.1145/2736277.2741669.

Jiang, J., W. Jeng, and D. He. (2013). "How Do Users Respond to Voice Input Errors? Lexical and Phonetic Query Reformulation in Voice Search". In: *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 143–152. DOI: 10.1145/2484028.2484092.

Joachims, T., L. Granka, B. Pan, H. Hembrooke, and G. Gay. (2005). "Accurately Interpreting Clickthrough Data as Implicit Feedback". In: *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 154–161. DOI: 10.1145/1076034.1076063.

Joko, H. and F. Hasibi. (2022). "Personal Entity, Concept, and Named Entity Linking in Conversations". In: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 4099–4103. DOI: 10.1145/3511808.3557667.

Joko, H., F. Hasibi, K. Balog, and A. P. de Vries. (2021). "Conversational Entity Linking: Problem Definition and Datasets". In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2390–2397. DOI: 10.1145/3404835.3463258.

Jones, M., G. Marsden, N. Mohd-Nasir, K. Boone, and G. Buchanan. (1999). "Improving Web Interaction on Small Displays". *Comput. Netw.* 31(11–16): 1129–1137. DOI: 10.1016/S1389-1286(99)00013-4.

Ju, Y., F. Zhao, S. Chen, B. Zheng, X. Yang, and Y. Liu. (2019). "Technical report on Conversational Question Answering". *ArXiv*. abs/1909.10772.

Jurafsky, D. and J. H. Martin. (2021). *Speech and Language Processing (3rd Edition)*. Prentice-Hall, Inc.

Kacupaj, E., J. Plepi, K. Singh, H. Thakkar, J. Lehmann, and M. Maleshkova. (2021). "Conversational Question Answering over Knowledge Graphs with Transformer and Graph Attention Networks". In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. 850–862. DOI: 10.18653/v1/2021.eacl-main.72.

Kacupaj, E., K. Singh, M. Maleshkova, and J. Lehmann. (2022). "Contrastive Representation Learning for Conversational Question Answering over Knowledge Graphs". In: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 925–934. DOI: 10.1145/3511808.3557267.

Kaiser, M., R. Saha Roy, and G. Weikum. (2020). "Conversational Question Answering over Passages by Leveraging Word Proximity Networks": 2129–2132. DOI: 10.1145/3397271.3401399.

Kaiser, M., R. Saha Roy, and G. Weikum. (2021). "Reinforcement Learning from Reformulations in Conversational Question Answering over Knowledge Graphs". In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 459–469. DOI: 10.1145/3404835.3462859.

Kaisser, M., M. A. Hearst, and J. B. Lowe. (2008). "Improving search results quality by customizing summary lengths". In: *Proceedings of ACL-08: HLT*. 701–709.

Kanoulas, E., B. Carterette, P. D. Clough, and M. Sanderson. (2011). "Evaluating Multi-Query Sessions". In: *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1053–1062. DOI: 10.1145/2009916.2010056.

Kaushik, A., V. Bhat Ramachandra, and G. J. F. Jones. (2020). "An Interface for Agent Supported Conversational Search". In: *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. 452–456. DOI: 10.1145/3343413.3377942.

Keenoy, K. and M. Levene. (2003). "Personalisation of Web Search". In: *Proceedings of the 2003 International Conference on Intelligent Techniques for Web Personalization*. 201–228. DOI: 10.1007/11577935_11.

Kelly, D. (2009). “Methods for Evaluating Interactive Information Retrieval Systems with Users”. *Found. Trends Inf. Retr.* 3(1—2): 1–224. DOI: 10.1561/1500000012.

Kelly, D. and C. R. Sugimoto. (2013). “A systematic review of interactive information retrieval evaluation studies, 1967-2006”. *J. Assoc. Inf. Sci. Technol.* 64(4): 745–770.

Khattab, O., C. Potts, and M. Zaharia. (2021a). “Baleen: Robust Multi-Hop Reasoning at Scale via Condensed Retrieval”. In: *Advances in Neural Information Processing Systems.* Ed. by A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan. URL: https://openreview.net/forum?id=Ghk0AJ8XtVx.

Khattab, O., C. Potts, and M. Zaharia. (2021b). “Relevance-guided Supervision for OpenQA with ColBERT”. *Transactions of the Association for Computational Linguistics.* 9: 929–944. DOI: 10.1162/tacl_a_00405.

Kiesel, J., L. Meyer, M. Potthast, and B. Stein. (2021a). “Meta-Information in Conversational Search”. *ACM Transactions on Information Systems (ACM TOIS).* 39(4). Ed. by C. Hauff, J. Kiseleva, M. Sanderson, H. Zamani, and Y. Zhang. DOI: 10.1145/3468868.

Kiesel, J., D. Spina, H. Wachsmuth, and B. Stein. (2021b). “The Meant, the Said, and the Understood: Conversational Argument Search and Cognitive Biases”. In: *CUI 2021 - 3rd Conference on Conversational User Interfaces.* DOI: 10.1145/3469595.3469615.

Kim, G., H. Kim, J. Park, and J. Kang. (2021). “Learn to Resolve Conversational Dependency: A Consistency Training Framework for Conversational Question Answering”. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers).* 6130–6141. DOI: 10.18653/v1/2021.acl-long.478.

Kim, J., P. Thomas, R. Sankaranarayana, T. Gedeon, and H.-J. Yoon. (2015). “Eye-Tracking Analysis of User Behavior and Performance in Web Search on Large and Small Screens”. *J. Assoc. Inf. Sci. Technol.* 66(3): 526–544. DOI: 10.1002/asi.23187.

Kim, J., P. Thomas, R. Sankaranarayana, T. Gedeon, and H.-J. Yoon. (2017). "What Snippet Size is Needed in Mobile Web Search?" In: *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. 97–106. DOI: 10.1145/3020165.3020173.

Kim, S., M. Eric, K. Gopalakrishnan, B. Hedayatnia, Y. Liu, and D. Hakkani-Tur. (2020). "Beyond Domain APIs: Task-oriented Conversational Modeling with Unstructured Knowledge Access". In: *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 278–289. URL: https://aclanthology.org/2020.sigdial-1.35.

Kiseleva, J., K. Williams, A. Hassan Awadallah, A. C. Crook, I. Zitouni, and T. Anastasakos. (2016). "Predicting User Satisfaction with Intelligent Assistants". In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 45–54. DOI: 10.1145/2911451.2911521.

Kitano, H. and C. Van Ess-Dykema. (1991). "Toward a Plan-Based Understanding Model for Mixed-Initiative Dialogues". In: *Proceedings of the 29th Annual Meeting on Association for Computational Linguistics*. 25–32. DOI: 10.3115/981344.981348.

Kolomiyets, O. and M.-F. Moens. (2011). "A Survey on Question Answering Technology from an Information Retrieval Perspective". *Inf. Sci.* 181(24): 5412–5434. DOI: 10.1016/j.ins.2011.07.047.

Koman, J., K. Fauvelle, S. Schuck, N. Texier, and A. Mebarki. (2020). "Physicians' Perceptions of the Use of a Chatbot for Information Seeking: Qualitative Study". *Journal of medical Internet research*. 22(11): e15185.

Konstan, J. A. and J. Riedl. (2012). "Recommender Systems: From Algorithms to User Experience". *User Modeling and User-Adapted Interaction*. 22(1–2): 101–123. DOI: 10.1007/s11257-011-9112-x.

Kostric, I., K. Balog, and F. Radlinski. (2021). "Soliciting User Preferences in Conversational Recommender Systems via Usage-Related Questions". In: *Proceedings of the 15th ACM Conference on Recommender Systems*. 724–729. DOI: 10.1145/3460231.3478861.

Krasakis, A. M., A. Yates, and E. Kanoulas. (2022). "Zero-Shot Query Contextualization for Conversational Search". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1880–1884. DOI: 10.1145/3477495.3531769.

Krebs, L. M., O. L. Alvarado Rodriguez, P. Dewitte, J. Ausloos, D. Geerts, L. Naudts, and K. Verbert. (2019). "Tell Me What You Know: GDPR Implications on Designing Transparency and Accountability for News Recommender Systems". In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6. DOI: 10.1145/3290607.3312808.

Krishna, K., A. Roy, and M. Iyyer. (2021). "Hurdles to Progress in Long-form Question Answering". In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 4940–4957. DOI: 10.18653/v1/2021.naacl-main.393.

Krum, U., H. Holzapfel, and A. Waibel. (2005). "Clarification questions to improve dialogue flow and speech recognition in spoken dialogue systems". In: *INTERSPEECH 2005 - Eurospeech, 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, September 4-8, 2005*. 3417–3420.

Ku, A., P. Anderson, R. Patel, E. Ie, and J. Baldridge. (2020). "Room-Across-Room: Multilingual Vision-and-Language Navigation with Dense Spatiotemporal Grounding". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 4392–4412. DOI: 10.18653/v1/2020.emnlp-main.356.

Kumar, V. and J. Callan. (2020). "Making Information Seeking Easier: An Improved Pipeline for Conversational Search". In: *Findings of the Association for Computational Linguistics: EMNLP 2020*. 3971–3980. DOI: 10.18653/v1/2020.findings-emnlp.354.

Kumar, V., V. Raunak, and J. Callan. (2020). "Ranking Clarification Questions via Natural Language Inference". In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2093–2096. DOI: 10.1145/3340531.3412137.

Kung, T. H., M. Cheatham, A. Medinilla, ChatGPT, C. Sillos, L. A. D. Leon, C. Elepano, M. A. Madriaga, R. G. Aggabao, G. Diaz-Candido, J. M. Z. Maningo, and V. Tseng. (2022). "Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models". In: *medRxiv*.

Laban, G. and T. Araujo. (2020). "The Effect of Personalization Techniques in Users' Perceptions of Conversational Recommender Systems". In: *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. DOI: 10.1145/3383652.3423890.

Lai, J., C. Karat, and N. Yankelovich. (2009). "Conversational speech interfaces and technologies". *Human-Computer Interaction: Design Issues, Solutions, and Applications*: 53–63.

Lan, Z., M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut. (2020). "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations". In: *International Conference on Learning Representations*. URL: https://openreview.net/forum?id=H1eA7AEtvS.

Lei, W., X. He, Y. Miao, Q. Wu, R. Hong, M.-Y. Kan, and T.-S. Chua. (2020a). "Estimation-Action-Reflection: Towards Deep Interaction Between Conversational and Recommender Systems". In: *Proceedings of the 13th International Conference on Web Search and Data Mining*. 304–312. DOI: 10.1145/3336191.3371769.

Lei, W., G. Zhang, X. He, Y. Miao, X. Wang, L. Chen, and T.-S. Chua. (2020b). "Interactive Path Reasoning on Graph for Conversational Recommendation". In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2073–2083. DOI: 10.1145/3394486.3403258.

Leszczynski, M. E., R. Ganti, S. Zhang, K. Balog, F. Radlinski, F. Pereira, and A. T. Chaganty. (2022). "Conversational Music Retrieval with Synthetic Data". In: *Second Workshop on Interactive Learning for Natural Language Processing at NeurIPS 2022*.

Lewis, P., E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela. (2020). "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks". In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*.

Li, R., S. E. Kahou, H. Schulz, V. Michalski, L. Charlin, and C. Pal. (2018). "Towards Deep Conversational Recommendations". In: *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*.

Li, S., W. Lei, Q. Wu, X. He, P. Jiang, and T.-S. Chua. (2021). "Seamlessly Unifying Attributes and Items: Conversational Recommendation for Cold-Start Users". *ACM Trans. Inf. Syst.* 39(4). DOI: 10.1145/3446427.

Li, S., R. Xie, Y. Zhu, X. Ao, F. Zhuang, and Q. He. (2022a). "User-Centric Conversational Recommendation with Multi-Aspect User Modeling". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 223–233. DOI: 10.1145/3477495.3532074.

Li, S., B. P. Majumder, and J. McAuley. (2022b). "Self-Supervised Bot Play for Transcript-Free Conversational Recommendation with Rationales". In: *Proceedings of the 16th ACM Conference on Recommender Systems*. 327–337. DOI: 10.1145/3523227.3546783.

Liao, L., L. H. Long, Z. Zhang, M. Huang, and T.-S. Chua. (2021). "MMConv: An Environment for Multimodal Conversational Search across Multiple Domains". In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 675–684. DOI: 10.1145/3404835.3462970.

Lim, W. M., S. Kumar, S. Verma, and R. Chaturvedi. (2022). "Alexa, what do we know about conversational commerce? Insights from a systematic literature review". *Psychology & Marketing*. 39(6): 1129–1155.

Lin, A., J. Wang, Z. Zhu, and J. Caverlee. (2022). "Quantifying and Mitigating Popularity Bias in Conversational Recommender Systems". In: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 1238–1247. DOI: 10.1145/3511808.3557423.

Lin, C.-Y. (2004). "ROUGE: A Package for Automatic Evaluation of Summaries". In: *Text Summarization Branches Out*. 74–81. URL: https://www.aclweb.org/anthology/W04-1013.

Lin, J., R. Nogueira, and A. Yates. (2020a). "Pretrained Transformers for Text Ranking: BERT and Beyond". *CoRR*. abs/2010.06467. URL: https://arxiv.org/abs/2010.06467.

Lin, S.-C., J.-H. Yang, and J. Lin. (2021a). "Contextualized Query Embeddings for Conversational Search". In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 1004–1015. DOI: 10.18653/v1/2021.emnlp-main.77.

Lin, S.-C., J.-H. Yang, R. Nogueira, M.-F. Tsai, C.-J. Wang, and J. J. Lin. (2020b). "Query Reformulation using Query History for Passage Retrieval in Conversational Search". *ArXiv*. abs/2005.02230.

Lin, S.-C., J.-H. Yang, R. Nogueira, M.-F. Tsai, C.-J. Wang, and J. J. Lin. (2021b). "Multi-Stage Conversational Passage Retrieval: An Approach to Fusing Term Importance Estimation and Neural Query Rewriting". *ACM Transactions on Information Systems (TOIS)*. 39: 1–29.

Liono, J., M. S. Rahaman, F. D. Salim, Y. Ren, D. Spina, F. Scholer, J. R. Trippas, M. Sanderson, P. N. Bennett, and R. W. White. (2020). "Intelligent Task Recognition: Towards Enabling Productivity Assistance in Daily Life". In: *International Conference on Multimedia Retrieval (ICMR'20)*.

Lipani, A., B. Carterette, and E. Yilmaz. (2019). "From a User Model for Query Sessions to Session Rank Biased Precision (SRBP)". In: *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*. 109–116. DOI: 10.1145/3341981.3344216.

Lipani, A., B. Carterette, and E. Yilmaz. (2021). "How Am I Doing?: Evaluating Conversational Search Systems Offline". *ACM Trans. Inf. Syst.* 39(4). DOI: 10.1145/3451160.

Liu, C.-W., R. Lowe, I. Serban, M. Noseworthy, L. Charlin, and J. Pineau. (2016). "How NOT To Evaluate Your Dialogue System: An Empirical Study of Unsupervised Evaluation Metrics for Dialogue Response Generation". In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 2122–2132. DOI: 10.18653/v1/D16-1230.

Liu, Z., K. Zhou, and M. L. Wilson. (2021a). "Meta-Evaluation of Conversational Search Evaluation Metrics". *ACM Trans. Inf. Syst.* 39(4). DOI: 10.1145/3445029.

Liu, Z., P. Ren, Z. Chen, Z. Ren, M. de Rijke, and M. Zhou. (2021b). "Learning to Ask Conversational Questions by Optimizing Levenshtein Distance". In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 5638–5650. DOI: 10.18653/v1/2021.acl-long.438.

Louis, A., F. Radlinski, and D. Roth. (2020). ""I'd rather just go to bed": Understanding Indirect Answers". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. DOI: 10.18653/v1/2020.emnlp-main.601.

Lowe, R., N. Pow, I. Serban, and J. Pineau. (2015). "The Ubuntu Dialogue Corpus: A Large Dataset for Research in Unstructured Multi-Turn Dialogue Systems". In: *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 285–294. DOI: 10.18653/v1/W15-4640.

Lu, X., S. Pramanik, R. Saha Roy, A. Abujabal, Y. Wang, and G. Weikum. (2019). "Answering Complex Questions by Joining Multi-Document Evidence with Quasi Knowledge Graphs". In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 105–114. DOI: 10.1145/3331184.3331252.

Ma, Q., O. Bojar, and Y. Graham. (2018). "Results of the WMT18 Metrics Shared Task: Both characters and embeddings achieve good performance". In: *Proceedings of the Third Conference on Machine Translation: Shared Task Papers*. 671–688. DOI: 10.18653/v1/W18-6450.

Mahmood, T. and F. Ricci. (2009). "Improving Recommender Systems with Adaptive Conversational Strategies". In: *Proceedings of the 20th ACM Conference on Hypertext and Hypermedia*. 73–82. DOI: 10.1145/1557914.1557930.

Mallinson, J., A. Severyn, E. Malmi, and G. Garrido. (2020). "FELIX: Flexible Text Editing Through Tagging and Insertion". In: *Findings of the Association for Computational Linguistics: EMNLP 2020*. 1244–1255. DOI: 10.18653/v1/2020.findings-emnlp.111.

Malmi, E., S. Krause, S. Rothe, D. Mirylenka, and A. Severyn. (2019). "Encode, Tag, Realize: High-Precision Text Editing". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 5054–5065. DOI: 10.18653/v1/D19-1510.

Mandya, A., J. O' Neill, D. Bollegala, and F. Coenen. (2020). "Do not let the history haunt you: Mitigating Compounding Errors in Conversational Question Answering". English. In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*. 2017–2025. URL: https://aclanthology.org/2020.lrec-1.248.

Mao, K., Z. Dou, and H. Qian. (2022). "Curriculum Contrastive Context Denoising for Few-Shot Conversational Dense Retrieval". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 176–186. DOI: 10.1145/3477495.3531961.

Marchionini, G. (2006). "Exploratory Search: From Finding to Understanding". *Commun. ACM*. 49(4): 41–46. DOI: 10.1145/1121949.1121979.

Marion, P., P. Nowak, and F. Piccinno. (2021). "Structured Context and High-Coverage Grammar for Conversational Question Answering over Knowledge Graphs". In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 8813–8829. DOI: 10.18653/v1/2021.emnlp-main.695.

Matteson, M. L., J. Salamon, and L. Brewster. (2011). "A Systematic Review of Research on Live Chat Service". *Reference & User Services Quarterly*. 51(2): 172–190. URL: http://www.jstor.org/stable/refuseserq.51.2.172.

Maxwell, D., L. Azzopardi, and Y. Moshfeghi. (2017). “A Study of Snippet Length and Informativeness: Behaviour, Performance and User Experience”. In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 135–144. DOI: 10.1145/3077136.3080824.

McTear, M., Z. Callejas, and D. Griol. (2016). *The Conversational Interface: Talking to Smart Devices*. Springer.

McTear, M. F. (2017). “The Rise of the Conversational Interface: A New Kid on the Block?” In: *Future and Emerging Trends in Language Technology. Machine Learning and Big Data*. Ed. by J. F. Quesada, F.-J. Martín Mateos, and T. López Soto. 38–49.

Mehri, S., M. Eric, and D. Hakkani-Tur. (2020). “DialoGLUE: A Natural Language Understanding Benchmark for Task-Oriented Dialogue”. *ArXiv*. abs/2009.13570.

Mele, I., C. I. Muntean, F. M. Nardini, R. Perego, N. Tonellotto, and O. Frieder. (2020). “Topic Propagation in Conversational Search”. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2057–2060. DOI: 10.1145/3397271.3401268.

Mirzadeh, N., F. Ricci, and M. Bansal. (2005). “Feature selection methods for conversational recommender systems”. In: *2005 IEEE International Conference on e-Technology, e-Commerce and e-Service*. 772–777. DOI: 10.1109/EEE.2005.75.

Mitra, B. and N. Craswell. (2018). “An Introduction to Neural Information Retrieval”. *Foundations and Trends® in Information Retrieval*. 13(1): 1–126. DOI: 10.1561/1500000061.

Mitra, B., S. Hofstätter, H. Zamani, and N. Craswell. (2021). “Improving Transformer-Kernel Ranking Model Using Conformer and Query Term Independence”. In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1697–1702. DOI: 10.1145/3404835.3463049.

Montazeralghaem, A., J. Allan, and P. S. Thomas. (2021). “Large-Scale Interactive Conversational Recommendation System Using Actor-Critic Framework”. In: *Proceedings of the 15th ACM Conference on Recommender Systems*. 220–229. DOI: 10.1145/3460231.3474271.

Moon, S., P. Shah, A. Kumar, and R. Subba. (2019). "OpenDialKG: Explainable Conversational Reasoning with Attention-based Walks over Knowledge Graphs". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 845–854. DOI: 10.18653/v1/P19-1081.

Mori, M., K. F. MacDorman, and N. Kageki. (2012). "The Uncanny Valley [From the Field]". *IEEE Robotics Automation Magazine*. 19(2): 98–100. DOI: 10.1109/MRA.2012.2192811.

Morris, D., M. Ringel Morris, and G. Venolia. (2008). "SearchBar: A Search-Centric Web History for Task Resumption and Information Re-Finding". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1207–1216. DOI: 10.1145/1357054.1357242.

Müller, M. and M. Volk. (2013). "Statistical Machine Translation of Subtitles: From OpenSubtitles to TED". In: *GSCL*.

"Multimodal Interfaces". (2008). In: *Encyclopedia of Multimedia*. Ed. by B. Furht. Springer US. 651–652. DOI: 10.1007/978-0-387-78414-4_159.

Murray, G. C. and J. Teevan. (2007). "Query Log Analysis: Social and Technological Challenges". *SIGIR Forum*. 41(2): 112–120. DOI: 10.1145/1328964.1328985.

Myers, C., A. Furqan, J. Nebolsky, K. Caro, and J. Zhu. (2018). "Patterns for How Users Overcome Obstacles in Voice User Interfaces". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–7. URL: https://doi.org/10.1145/3173574.3173580.

Nag, P. and Ö. N. Yalçın. (2020). "Gender Stereotypes in Virtual Agents". In: *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. DOI: 10.1145/3383652.3423876.

Nass, C. and S. Brave. (2005). *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. The MIT Press.

Nie, L., W. Wang, R. Hong, M. Wang, and Q. Tian. (2019). "Multimodal Dialog System: Generating Responses via Adaptive Decoders". In: *Proceedings of the 27th ACM International Conference on Multimedia*. 1098–1106. DOI: 10.1145/3343031.3350923.

Novick, D. G. and S. A. Douglas. (1988). "Control of Mixed-Initiative Discourse through Meta-Locutionary Acts: A Computational Model". *Tech. rep.*

Oard, D. W. and J. Kim. (1998). "Implicit Feedback for Recommender Systems". In: *Proceedings of the AAAI Workshop on Recommender Systems*. AAAI.

Oddy, R. N. (1977). "Information retrieval through man-machine dialogue". *Journal of Documentation*. 33(1): 1–14.

Ohsugi, Y., I. Saito, K. Nishida, H. Asano, and J. Tomita. (2019). "A Simple but Effective Method to Incorporate Multi-turn Context with BERT for Conversational Machine Comprehension". In: *Proceedings of the First Workshop on NLP for Conversational AI*. 11–17. DOI: 10.18653/v1/W19-4102.

Ong, K., K. Järvelin, M. Sanderson, and F. Scholer. (2017). "Using Information Scent to Understand Mobile and Desktop Web Search Behavior". In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 295–304. DOI: 10.1145/3077136.3080817.

OpenAI. (2023). "GPT-4 Technical Report". arXiv: 2303.08774 [cs.CL].

Oraby, S., P. Gundecha, J. Mahmud, M. Bhuiyan, and R. Akkiraju. (2017). ""How May I Help You?": Modeling Twitter Customer ServiceConversations Using Fine-Grained Dialogue Acts". In: *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. 343–355. DOI: 10.1145/3025171.3025191.

Ouyang, S., Z. Zhang, and H. Zhao. (2021). "Dialogue Graph Modeling for Conversational Machine Reading". In: *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. 3158–3169. DOI: 10.18653/v1/2021.findings-acl.279.

Oviatt, S. and P. R. Cohen. (2015). *The Paradigm Shift to Multimodality in Contemporary Computer Interfaces*. Morgan & Claypool Publishers.

Owoicho, P., J. Dalton, M. Aliannejadi, L. Azzopardi, J. R. Trippas, and S. Vakulenko. (2022). "TREC CAsT 2022: Going Beyond User Ask and System Retrieve with Initiative and Response Generation". In: *TREC*.

Owoicho, P., J. Dalton, M. Aliannejadi, L. Azzopardi, J. R. Trippas, and S. Vakulenko. (2023). "TREC CAsT 2022: Going Beyond User Ask and System Retrieve with Initiative and Response Generation". In: *Proceedings of the NIST Text Retrieval Conference (TREC 2022)*. 1–11.

Paek, T., S. Dumais, and R. Logan. (2004). "WaveLens: A New View onto Internet Search Results". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 727–734. DOI: 10.1145/985692.985784.

Pajukoski, J. (2018). "Impact of chat layout on usability in customer service chat multitasking". *MA thesis*. Aalto University. 85.

Papenmeier, A., A. Frummet, and D. Kern. (2022). ""Mhm..." – Conversational Strategies For Product Search Assistants". In: *ACM SIGIR Conference on Human Information Interaction and Retrieval*. 36–46. DOI: 10.1145/3498366.3505809.

Papenmeier, A., D. Kern, D. Hienert, A. Sliwa, A. Aker, and N. Fuhr. (2021). "Dataset of Natural Language Queries for E-Commerce". In: *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. 307–311. DOI: 10.1145/3406522.3446043.

Papineni, K., S. Roukos, T. Ward, and W.-J. Zhu. (2002). "Bleu: a Method for Automatic Evaluation of Machine Translation". In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. 311–318. DOI: 10.3115/1073083.1073135.

Paraschakis, D. (2016). "Recommender Systems from an Industrial and Ethical Perspective". In: *Proceedings of the 10th ACM Conference on Recommender Systems*. 463–466. DOI: 10.1145/2959100.2959101.

Parikh, A., O. Täckström, D. Das, and J. Uszkoreit. (2016). "A Decomposable Attention Model for Natural Language Inference". In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 2249–2255. DOI: 10.18653/v1/D16-1244.

Park, D., H. Yuan, D. Kim, Y. Zhang, M. Spyros, Y.-B. Kim, R. Sarikaya, E. Guo, Y. Ling, K. Quinn, P. Hung, B. Yao, and S. Lee. (2020). "Large-scale Hybrid Approach for Predicting User Satisfaction with Conversational Agents". arXiv: 2006.07113.

Peckham, J. (1991). “Speech Understanding and Dialogue over the telephone: an overview of the ESPRIT SUNDIAL project.” In: *Speech and Natural Language: Proceedings of a Workshop Held at Pacific Grove, California, February 19-22, 1991*.

Penha, G., A. Bălan, and C. Hauff. (2019). “Introducing MANtIS: a novel Multi-Domain Information Seeking Dialogues Dataset”. *ArXiv*. abs/1912.04639.

Penha, G. and C. Hauff. (2020). “What Does BERT Know about Books, Movies and Music? Probing BERT for Conversational Recommendation”. In: *Proceedings of the 14th ACM Conference on Recommender Systems*. 388–397. DOI: 10.1145/3383313.3412249.

Penha, G., S. Vakulenko, O. Dusek, L. Clark, V. Pal, and V. Adlakha. (2022). “The Seventh Workshop on Search-Oriented Conversational Artificial Intelligence (SCAI’22)”. In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 3466–3469. DOI: 10.1145/3477495.3531700.

Peshterliev, S., B. Oguz, D. Chatterjee, H. Inan, and V. Bhardwaj. (2021). “Conversational Answer Generation and Factuality for Reading Comprehension Question-Answering”. URL: https://arxiv.org/abs/2103.06500.

Peters, M. E., M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. (2018). “Deep Contextualized Word Representations”. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. 2227–2237. DOI: 10.18653/v1/N18-1202.

Plepi, J., E. Kacupaj, K. Singh, H. Thakkar, and J. Lehmann. (2021). “Context Transformer with Stacked Pointer Networks for Conversational Question Answering over Knowledge Graphs”. In: *Eighteenth Extended Semantic Web Conference - Research Track*. URL: https://openreview.net/forum?id=Ecd_zT8KJoQ.

Pommeranz, A., J. Broekens, P. Wiggers, W.-P. Brinkman, and C. M. Jonker. (2012). “Designing Interfaces for Explicit Preference Elicitation: A User-Centered Investigation of Preference Representation and Elicitation Process”. *User Modeling and User-Adapted Interaction*. 22(4–5): 357–397. DOI: 10.1007/s11257-011-9116-6.

Ponte, J. M. and W. B. Croft. (1998). “A Language Modeling Approach to Information Retrieval”. In: *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 275–281. DOI: 10.1145/290941.291008.

Pradeep, R., X. Ma, R. Nogueira, and J. Lin. (2021). *Vera: Prediction Techniques for Reducing Harmful Misinformation in Consumer Health Search*. DOI: 10.1145/3404835.3463120.

Prakash, P., J. Killingback, and H. Zamani. (2021). “Learning Robust Dense Retrieval Models from Incomplete Relevance Labels”. In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1728–1732. DOI: 10.1145/3404835.3463106.

Qu, C., L. Yang, C. Chen, W. B. Croft, K. Krishna, and M. Iyyer. (2021). “Weakly-Supervised Open-Retrieval Conversational Question Answering”. In: *Advances in Information Retrieval: 43rd European Conference on IR Research*. 529–543. DOI: 10.1007/978-3-030-72113-8_35.

Qu, C., L. Yang, C. Chen, M. Qiu, W. B. Croft, and M. Iyyer. (2020). “Open-Retrieval Conversational Question Answering”. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 539–548. DOI: 10.1145/3397271.3401110.

Qu, C., L. Yang, W. B. Croft, J. R. Trippas, Y. Zhang, and M. Qiu. (2018). “Analyzing and Characterizing User Intent in Information-seeking Conversations”. In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 989–992. DOI: 10.1145/3209978.3210124.

Qu, C., L. Yang, W. B. Croft, Y. Zhang, J. R. Trippas, and M. Qiu. (2019a). “User Intent Prediction in Information-seeking Conversations”. In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. 25–33. DOI: 10.1145/3295750.3298924.

Qu, C., L. Yang, M. Qiu, W. B. Croft, Y. Zhang, and M. Iyyer. (2019b). “BERT with History Answer Embedding for Conversational Question Answering”. In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1133–1136. DOI: 10.1145/3331184.3331341.

Qu, C., L. Yang, M. Qiu, Y. Zhang, C. Chen, W. B. Croft, and M. Iyyer. (2019c). "Attentive History Selection for Conversational Question Answering". In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 1391–1400. DOI: 10.1145/3357384.3357905.

Radlinski, F., K. Balog, B. Byrne, and K. Krishnamoorthi. (2019). "Coached Conversational Preference Elicitation: A Case Study in Understanding Movie Preferences". In: *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*. 353–360. DOI: 10.18653/v1/W19-5941.

Radlinski, F., K. Balog, F. Diaz, L. G. Dixon, and B. Wedin. (2022a). "On Natural Language User Profiles for Transparent and Scrutable Recommendation". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22)*. DOI: 10.1145/3477495.3531873.

Radlinski, F., C. Boutilier, D. Ramachandran, and I. Vendrov. (2022b). "Subjective Attributes in Conversational Recommendation Systems: Challenges and Opportunities". In: *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI-22)*. 12287–12293. DOI: 10.1609/aaai.v36i11.21492.

Radlinski, F. and N. Craswell. (2017). "A Theoretical Framework for Conversational Search". In: *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. 117–126. DOI: 10.1145/3020165.3020183.

Rae, J. W., A. Potapenko, S. M. Jayakumar, C. Hillier, and T. P. Lillicrap. (2020). "Compressive Transformers for Long-Range Sequence Modelling". In: *International Conference on Learning Representations*. URL: https://openreview.net/forum?id=SylKikSYDH.

Raffel, C., N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu. (2020). "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer". *J. Mach. Learn. Res.* 21: 140:1–140:67.

Ram, A., R. Prasad, C. Khatri, A. Venkatesh, R. Gabriel, Q. Liu, J. Nunn, B. Hedayatnia, M. Cheng, A. Nagar, E. King, K. Bland, A. Wartick, Y. Pan, H. Song, S. Jayadevan, G. Hwang, and A. Pettigrue. (2018). "Conversational AI: The Science Behind the Alexa Prize". *ArXiv*. abs/1801.03604.

Ranzato, M., S. Chopra, M. Auli, and W. Zaremba. (2016). "Sequence Level Training with Recurrent Neural Networks". In: *Proceedings of the 4th International Conference on Learning Representations*. Ed. by Y. Bengio and Y. LeCun.

Rao, S. and H. Daumé III. (2018). "Learning to Ask Good Questions: Ranking Clarification Questions using Neural Expected Value of Perfect Information". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2737–2746. URL: https://www.aclweb.org/anthology/P18-1255.

Rao, S. and H. Daumé III. (2019). "Answer-based Adversarial Training for Generating Clarification Questions". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 143–155. URL: https://www.aclweb.org/anthology/N19-1013.

Rashkin, H., V. Nikolaev, M. Lamm, M. Collins, D. Das, S. Petrov, G. S. Tomar, I. Turc, and D. Reitter. (2021). "Measuring Attribution in Natural Language Generation Models". arXiv: 2112.12870 [cs.CL].

Rastogi, A., X. Zang, S. Sunkara, R. Gupta, and P. Khaitan. (2020). "Schema-Guided Dialogue State Tracking Task at DSTC8". *ArXiv*. abs/2002.01359.

Rastogi, P., A. Gupta, T. Chen, and M. Lambert. (2019). "Scaling Multi-Domain Dialogue State Tracking via Query Reformulation". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Industry Papers)*. 97–105. DOI: 10.18653/v1/N19-2013.

Reddy, S., D. Chen, and C. D. Manning. (2019). "CoQA: A Conversational Question Answering Challenge". *Transactions of the Association for Computational Linguistics*. 7(Mar.): 249–266. DOI: 10.1162/tacl_a_00266.

Reichman, R. (1985). *Getting Computers to Talk Like You and Me: Discourse Context, Focus, and Semantics:(an ATN Model)*. MIT press.

Ren, P., Z. Chen, Z. Ren, E. Kanoulas, C. Monz, and M. De Rijke. (2021a). "Conversations with Search Engines: SERP-Based Conversational Response Generation". *ACM Trans. Inf. Syst.* 39(4). DOI: 10.1145/3432726.

Ren, P., Z. Liu, X. Song, H. Tian, Z. Chen, Z. Ren, and M. de Rijke. (2021b). "Wizard of Search Engine: Access to Information Through Conversations with Search Engines". In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 533–543. DOI: 10.1145/3404835.3462897.

Ren, X., H. Yin, T. Chen, H. Wang, N. Q. V. Hung, Z. Huang, and X. Zhang. (2020). "CRSAL: Conversational Recommender Systems with Adversarial Learning". *ACM Trans. Inf. Syst.* 38(4). DOI: 10.1145/3394592.

Resnick, P., N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. (1994). "GroupLens: An Open Architecture for Collaborative Filtering of Netnews". In: *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*. 175–186. DOI: 10.1145/192844.192905.

Resnick, P. and H. R. Varian. (1997). "Recommender Systems". *Commun. ACM*. 40(3): 56–58. DOI: 10.1145/245108.245121.

Ricci, F., L. Rokach, B. Shapira, and P. B. Kantor. (2010). *Recommender Systems Handbook*. 1st.

Rieser, V. and J. Moore. (2005). "Implications for Generating Clarification Requests in Task-Oriented Dialogues". In: *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*. 239–246. DOI: 10.3115/1219840.1219870.

Roller, S., E. Dinan, N. Goyal, D. Ju, M. Williamson, Y. Liu, J. Xu, M. Ott, E. M. Smith, Y.-L. Boureau, and J. Weston. (2021). "Recipes for Building an Open-Domain Chatbot". In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. 300–325. DOI: 10.18653/v1/2021.eacl-main.24.

Rose, D. E., D. Orr, and R. G. P. Kantamneni. (2007). "Summary Attributes and Perceived Search Quality". In: *Proceedings of the 16th International Conference on World Wide Web*. 1201–1202. DOI: 10.1145/1242572.1242765.

Rosset, C., C. Xiong, X. Song, D. Campos, N. Craswell, S. Tiwary, and P. Bennett. (2020). "Leading Conversational Search by Suggesting Useful Questions". In: *Proceedings of The Web Conference 2020*. 1160–1170. DOI: 10.1145/3366423.3380193.

Rothe, S., S. Narayan, and A. Severyn. (2020). "Leveraging Pre-trained Checkpoints for Sequence Generation Tasks". *Transactions of the Association for Computational Linguistics*. 8: 264–280. DOI: 10.1162/tacl_a_00313.

Rousseau, C., Y. Bellik, F. Vernier, and D. Bazalgette. (2006). "A Framework for the Intelligent Multimodal Presentation of Information". *Signal Process.* 86(12): 3696–3713. DOI: 10.1016/j.sigpro.2006.02.041.

Rudnicky, A. I. (2005). "Multimodal Dialogue Systems". In: *Spoken Multimodal Human-Computer Dialogue in Mobile Environments*. Ed. by W. Minker, D. Bühler, and L. Dybkjær. Springer Netherlands. 3–11. DOI: 10.1007/1-4020-3075-4_1.

Sabei, I., A. Mourad, and G. Zuccon. (2022). "SCC - A Test Collection for Search in Chat Conversations". In: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 4429–4433. DOI: 10.1145/3511808.3557692.

Sachse, J. (2019). "The influence of snippet length on user behavior in mobile web search". *Aslib Journal of Information Management*.

Saeidi, M., M. Bartolo, P. Lewis, S. Singh, T. Rocktäschel, M. Sheldon, G. Bouchard, and S. Riedel. (2018). "Interpretation of Natural Language Rules in Conversational Machine Reading". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2087–2097. DOI: 10.18653/v1/D18-1233.

Saha, A., V. Pahuja, M. M. Khapra, K. Sankaranarayanan, and S. Chandar. (2018). "Complex Sequential Question Answering: Towards Learning to Converse over Linked Question Answer Pairs with a Knowledge Graph". In: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*.

Salle, A., S. Malmasi, O. Rokhlenko, and E. Agichtein. (2021). "Studying the Effectiveness of Conversational Search Refinement Through User Simulation". In: *Advances in Information Retrieval (ECIR 2021)*. URL: https://doi.org/10.1007/978-3-030-72113-8_39.

Salton, G. (1970). "Evaluation problems in interactive information retrieval". *Information Storage and Retrieval*. 6(1): 29–44. DOI: https://doi.org/10.1016/0020-0271(70)90011-2.

Schaffer, S. and N. Reithinger. (2019). "Conversation is Multimodal: Thus Conversational User Interfaces Should Be as Well". In: *Proceedings of the 1st International Conference on Conversational User Interfaces*. DOI: 10.1145/3342775.3342801.

Schnabel, T., P. N. Bennett, S. T. Dumais, and T. Joachims. (2016). "Using Shortlists to Support Decision Making and Improve Recommender System Performance". In: *Proceedings of the 25th International Conference on World Wide Web*. 987–997. DOI: 10.1145/2872427.2883012.

See, A., P. J. Liu, and C. D. Manning. (2017). "Get To The Point: Summarization with Pointer-Generator Networks". In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1073–1083. DOI: 10.18653/v1/P17-1099.

Sekulić, I., M. Aliannejadi, and F. Crestani. (2022). "Evaluating Mixed-initiative Conversational Search Systems via User Simulation". In: *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining (WSDM '22)*. URL: https://doi.org/10.1145/3488560.3498440.

Sellam, T., D. Das, and A. Parikh. (2020). "BLEURT: Learning Robust Metrics for Text Generation". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 7881–7892. DOI: 10.18653/v1/2020.acl-main.704.

Seo, M., A. Kembhavi, A. Farhadi, and H. Hajishirzi. (2017). "Bidirectional Attention Flow for Machine Comprehension". In: *International Conference on Learning Representations*. URL: https://openreview.net/forum?id=HJ0UKP9ge.

Sepliarskaia, A., J. Kiseleva, F. Radlinski, and M. de Rijke. (2018). "Preference Elicitation as an Optimization Problem". In: *Proceedings of the 12th ACM Conference on Recommender Systems*. 172–180. DOI: 10.1145/3240323.3240352.

Serban, I. V., R. Lowe, P. Henderson, L. Charlin, and J. Pineau. (2018). "A survey of available corpora for building data-driven dialogue systems: The journal version". *Dialogue Discourse*. 9(1): 1–49.

Shen, T., X. Geng, T. Qin, D. Guo, D. Tang, N. Duan, G. Long, and D. Jiang. (2019). "Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 2442–2451. DOI: 10.18653/v1/D19-1248.

Shen, X., B. Tan, and C. Zhai. (2005). "Context-sensitive Information Retrieval Using Implicit Feedback". In: *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 43–50. DOI: 10.1145/1076034.1076045.

Shuster, K., S. Poff, M. Chen, D. Kiela, and J. Weston. (2021). "Retrieval Augmentation Reduces Hallucination in Conversation". In: *Findings of the Association for Computational Linguistics: EMNLP 2021*. 3784–3803. DOI: 10.18653/v1/2021.findings-emnlp.320.

Skantze, G. (2007). "Error Handling in Spoken Dialogue Systems-Managing Uncertainty, Grounding and Miscommunication". *PhD thesis*. KTH, Stockholm.

Slokom, M. (2018). “Comparing Recommender Systems Using Synthetic Data”. In: *Proceedings of the 12th ACM Conference on Recommender Systems*. 548–552. DOI: 10.1145/3240323.3240325.

So, D., Q. Le, and C. Liang. (2019). “The evolved transformer”. In: *International Conference on Machine Learning*. PMLR. 5877–5886.

Sordoni, A., Y. Bengio, H. Vahabi, C. Lioma, J. Grue Simonsen, and J.-Y. Nie. (2015). “A Hierarchical Recurrent Encoder-Decoder for Generative Context-Aware Query Suggestion”. In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. 553–562. DOI: 10.1145/2806416.2806493.

Speggiorin, A., J. Dalton, and A. Leuski. (2022). “TaskMAD: A Platform for Multimodal Task-Centric Knowledge-Grounded Conversational Experimentation”. In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 3240–3244. DOI: 10.1145/3477495.3531679.

Spina, D., J. R. Trippas, P. Thomas, H. Joho, K. Byström, L. Clark, N. Craswell, M. Czerwinski, D. Elsweiler, A. Frummet, S. Ghosh, J. Kiesel, I. Lopatovska, D. McDuff, S. Meyer, A. Mourad, P. Owoicho, S. P. Cherumanal, D. Russell, and L. Sitbon. (2021). “Report on the Future Conversations Workshop at CHIIR 2021”. *SIGIR Forum*. 55(1). DOI: 10.1145/3476415.3476421.

Staliūnaitė, I. and I. Iacobacci. (2020). “Compositional and Lexical Semantics in RoBERTa, BERT and DistilBERT: A Case Study on CoQA”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 7046–7056. DOI: 10.18653/v1/2020.emnlp-main.573.

Stephen, J., K. Collie, D. McLeod, A. Rojubally, K. Fergus, M. Speca, J. Turner, J. Taylor-Brown, S. Sellick, K. Burrus, and M. Elramly. (2014). “Talking with text: Communication in therapist-led, live chat cancer support groups”. *Social Science & Medicine*. 104: 178–186. DOI: https://doi.org/10.1016/j.socscimed.2013.12.001.

Stoyanchev, S., A. Liu, and J. Hirschberg. (2014). “Towards Natural Clarification Questions in Dialogue Systems”. In: *AISB ’14*. Vol. 20.

Su, H., X. Shen, R. Zhang, F. Sun, P. Hu, C. Niu, and J. Zhou. (2019). "Improving Multi-turn Dialogue Modelling with Utterance ReWriter". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 22–31. DOI: 10.18653/v1/P19-1003.

Su, Z., J. A. Schneider, and S. D. Young. (2021). "The Role of Conversational Agents for Substance Use Disorder in Social Distancing Contexts". *Substance Use & Misuse*. 56(11): 1732–1735.

Sun, Y. and Y. Zhang. (2018). "Conversational Recommender System". In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 235–244. DOI: 10.1145/3209978.3210002.

Sutskever, I., O. Vinyals, and Q. V. Le. (2014). "Sequence to Sequence Learning with Neural Networks". In: *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger. 3104–3112. URL: http://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf.

Sutton, S. and R. Cole. (1997). "The CSLU Toolkit: Rapid Prototyping of Spoken Language Systems". In: *Proceedings of the 10th Annual ACM Symposium on User Interface Software and Technology*. 85–86. DOI: 10.1145/263407.263517.

Tabassum, M., T. Kosiundefinedski, A. Frik, N. Malkin, P. Wijesekera, S. Egelman, and H. R. Lipford. (2019). "Investigating Users' Preferences and Expectations for Always-Listening Voice Assistants". *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3(4). DOI: 10.1145/3369807.

Tao, C., J. Feng, R. Yan, W. Wu, and D. Jiang. (2021). "A Survey on Response Selection for Retrieval-based Dialogues". In: *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*. Ed. by Z.-H. Zhou. 4619–4626. URL: https://doi.org/10.24963/ijcai.2021/627.

Tavakoli, L., J. R. Trippas, H. Zamani, F. Scholer, and M. Sanderson. (2022). "MIMICS-Duo: Offline & Online Evaluation of Search Clarification". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 3198–3208. DOI: 10.1145/3477495.3531750.

Tavakoli, L., H. Zamani, F. Scholer, W. B. Croft, and M. Sanderson. (2021). "Analyzing clarification in asynchronous information-seeking conversations". *Journal of the Association for Information Science and Technology*. DOI: https://doi.org/10.1002/asi.24562.

Teevan, J. (2020). "Conversational Search in the Enterprise". In: *Conversational Search (Dagstuhl Seminar 19461)*. Ed. by A. Anand, L. Cavedon, H. Joho, M. Sanderson, and B. Stein. Dagstuhl. 47.

Teevan, J., E. Adar, R. Jones, and M. A. S. Potts. (2007). "Information Re-Retrieval: Repeat Queries in Yahoo's Logs". In: *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 151–158. DOI: 10.1145/1277741.1277770.

Teevan, J., E. Cutrell, D. Fisher, S. M. Drucker, G. Ramos, P. André, and C. Hu. (2009). "Visual Snippets: Summarizing Web Pages for Search and Revisitation". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2023–2032. DOI: 10.1145/1518701.1519008.

Tenney, I., D. Das, and E. Pavlick. (2019). "BERT Rediscovers the Classical NLP Pipeline". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 4593–4601. DOI: 10.18653/v1/P19-1452.

Thomas, P., M. Czerwinski, D. McDuff, N. Craswell, and G. Mark. (2018). "Style and Alignment in Information-Seeking Conversation". In: *Proceedings of the 2018 Conference on Human Information Interaction and Retrieval*. 42–51. DOI: 10.1145/3176349.3176388.

Thomas, P., D. McDuff, M. Czerwinski, and N. Craswell. (2017). "MISC: A data set of information-seeking conversations". In: *Proceedings of the 1st International Workshop on Conversational Approaches to Information Retrieval*.

Thompson, C. A., M. H. Göker, and P. Langley. (2004). "A Personalized System for Conversational Recommendations". *J. Artif. Int. Res.* 21(1): 393–428.

Traum, D. and P. Heeman. (1996). "Utterance Units in Spoken Dialogue". In: *ECAI Workshop on Dialogue Processing in Spoken Language Systems*.

Traum, D. R. and S. Larsson. (2003). "The information state approach to dialogue management". In: *Current and New Directions in Discourse and Dialogue*. Springer. 325–353.

Trienes, J. and K. Balog. (2019). "Identifying Unclear Questions in Community Question Answering Websites". In: *Advances in Information Retrieval*. Ed. by L. Azzopardi, B. Stein, N. Fuhr, P. Mayr, C. Hauff, and D. Hiemstra. 276–289.

Trippas, J. R. (2019). "Spoken Conversational Search: Audio-only Interactive Information Retrieval". *PhD thesis*. RMIT, Melbourne.

Trippas, J. R., D. Spina, L. Cavedon, H. Joho, and M. Sanderson. (2018). "Informing the Design of Spoken Conversational Search: Perspective Paper". In: *Proceedings of the 2018 Conference on Human Information Interaction and Retrieval*. 32–41. DOI: 10.1145/3176349.3176387.

Trippas, J. R., D. Spina, L. Cavedon, and M. Sanderson. (2017). "How Do People Interact in Conversational Speech-Only Search Tasks: A Preliminary Analysis". In: *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. 325–328. DOI: 10.1145/3020165.3022144.

Trippas, J. R., D. Spina, M. Sanderson, and L. Cavedon. (2015a). "Results Presentation Methods for a Spoken Conversational Search System". In: *Proceedings of the First International Workshop on Novel Web Search Interfaces and Systems*. 13–15. DOI: 10.1145/2810355.2810356.

Trippas, J. R., D. Spina, M. Sanderson, and L. Cavedon. (2015b). "Towards Understanding the Impact of Length in Web Search Result Summaries over a Speech-only Communication Channel". In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 991–994. DOI: 10.1145/2766462.2767826.

Trippas, J. R., D. Spina, M. Sanderson, and L. Cavedon. (2021). "Accessing Media Via an Audio-Only Communication Channel: A Log Analysis". In: *CUI 2021 - 3rd Conference on Conversational User Interfaces*. DOI: 10.1145/3469595.3469623.

Trippas, J. R., D. Spina, F. Scholer, A. H. Awadallah, P. Bailey, P. N. Bennett, R. W. White, J. Liono, Y. Ren, F. D. Salim, and M. Sanderson. (2019). "Learning About Work Tasks to Inform Intelligent Assistant Design". In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. 5–14. DOI: 10.1145/3295750.3298934.

Trippas, J. R., D. Spina, P. Thomas, M. Sanderson, H. Joho, and L. Cavedon. (2020). "Towards a Model for Spoken Conversational Search". *Information Processing & Management*. 57(2): 102162.

Trippas, J. R. and P. Thomas. (2019). "Data Sets for Spoken Conversational Search". In: *Proceedings of the CHIIR 2019 Workshop on Barriers to Interactive IR Resources Re-use*. 14–18.

Turpin, A., Y. Tsegay, D. Hawking, and H. E. Williams. (2007). "Fast Generation of Result Snippets in Web Search". In: *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 127–134. DOI: 10.1145/1277741.1277766.

Vakulenko, S., E. Kanoulas, and M. De Rijke. (2021). "A Large-Scale Analysis of Mixed Initiative in Information-Seeking Dialogues for Conversational Search". *ACM Trans. Inf. Syst.* 39(4). DOI: 10.1145/3466796.

Vakulenko, S., S. Longpre, Z. Tu, and R. Anantha. (2020). "Question Rewriting for Conversational Question Answering". *ArXiv*. abs/2004.14652.

van Eeuwen, M. (2017). "Mobile conversational commerce: messenger chatbots as the next interface between businesses and consumers". URL: http://essay.utwente.nl/71706/.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. (2017). "Attention is All you Need". In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. 5998–6008. URL: http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf.

Volokhin, S., M. Collins, O. Rokhlenko, and E. Agichtein. (2022). "Generating and Validating Contextually Relevant Justifications for Conversational Recommendation". In: *ACM SIGIR Conference on Human Information Interaction and Retrieval*. 284–289. DOI: 10.1145/3498366.3505789.

Voorhees, E. (2004). "Overview of the TREC 2004 Question Answering Track". In: *TREC*.

Voorhees, E. (2005). "Overview of the TREC 2005 Question Answering Track". In: *TREC*.

Voorhees, E. M. *et al.* (1999). "The TREC-8 question answering track report". In: *Trec*. Vol. 99. 77–82.

Voskarides, N., D. Li, P. Ren, E. Kanoulas, and M. de Rijke. (2020). "Query Resolution for Conversational Search with Limited Supervision". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 921–930. DOI: 10.1145/3397271.3401130.

Vtyurina, A., C. L. A. Clarke, E. Law, J. R. Trippas, and H. Bota. (2020). "A Mixed-Method Analysis of Text and Audio Search Interfaces with Varying Task Complexity". In: *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*. 61–68. DOI: 10.1145/3409256.3409822.

Vtyurina, A., D. Savenkov, E. Agichtein, and C. L. A. Clarke. (2017). "Exploring Conversational Search With Humans, Assistants, and Wizards". In: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2187–2193. DOI: 10.1145/3027063.3053175.

Vuong, T., G. Jacucci, and T. Ruotsalo. (2018). "Naturalistic Digital Task Modeling for Personal Information Assistance via Continuous Screen Monitoring". In: *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. 778–785. DOI: 10.1145/3267305.3274130.

Wadhwa, S. and H. Zamani. (2021). "Towards System-Initiative Conversational Information Seeking". In: *Proceedings of the Second International Conference on Design of Experimental Search and Information Retrieval Systems*. 102–116.

Walker, M. and S. Whittaker. (1990). "Mixed Initiative in Dialogue: An Investigation into Discourse Segmentation". In: *Proceedings of the 28th Annual Meeting on Association for Computational Linguistics*. 70–78. DOI: 10.3115/981823.981833.

Walker, M. A., D. J. Litman, C. A. Kamm, and A. Abella. (1997). "PARADISE: A Framework for Evaluating Spoken Dialogue Agents". In: *35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*. 271–280. DOI: 10.3115/976909.979652.

Wang, Z. and Q. Ai. (2021). "Controlling the Risk of Conversational Search via Reinforcement Learning". In: *Proceedings of the Web Conference 2021*. 1968–1977. DOI: 10.1145/3442381.3449893.

Warren, D. H. D. and F. C. N. Pereira. (1982). "An Efficient Easily Adaptable System for Interpreting Natural Language Queries". *Comput. Linguist.* 8(3–4): 110–122.

Weeratunga, A. M., S. A. U. Jayawardana, P. M. A. K. Hasindu, W. P. M. Prashan, and S. Thelijjagoda. (2015). "Project Nethra - an intelligent assistant for the visually disabled to interact with internet services". In: *2015 IEEE 10th International Conference on Industrial and Information Systems (ICIIS)*. 55–59. DOI: 10.1109/ICIINFS.2015.7398985.

Weizenbaum, J. (1966). "ELIZA—a Computer Program for the Study of Natural Language Communication between Man and Machine". *Commun. ACM*. 9(1): 36–45. DOI: 10.1145/365153.365168.

Welleck, S., I. Kulikov, S. Roller, E. Dinan, K. Cho, and J. Weston. (2020). "Neural Text Generation With Unlikelihood Training". In: *International Conference on Learning Representations*.

Weston, J., E. Dinan, and A. H. Miller. (2018). "Retrieve and Refine: Improved Sequence Generation Models For Dialogue". In: *SCAI@EMNLP*.

White, R. W. and R. A. Roth. (2009). *Exploratory Search: Beyond the Query-Response Paradigm*. Morgan & Claypool Publishers.

Williams, J., A. Raux, and M. Henderson. (2016). "The Dialog State Tracking Challenge Series: A Review". *Dialogue Discourse*. 7: 4–33.

Williams, R. J. (1992). "Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning". *Mach. Learn.* 8(3–4): 229–256. DOI: 10.1007/BF00992696.

Wilson, T. D. (1999). "Models in information behaviour research". *Journal of Documentation*. 55(3): 249–270.

Winograd, T. (1972). "Understanding natural language". *Cognitive Psychology*. 3(1): 1–191. DOI: https://doi.org/10.1016/0010-0285(72)90002-3.

Winograd, T. (1974). "Five lectures on artificial intelligence". *Tech. rep.* Stanford University, computer science department.

Winters, R. M., N. Joshi, E. Cutrell, and M. R. Morris. (2019). "Strategies for Auditory Display of Social Media". *Ergonomics in Design*. 27(1): 11–15. DOI: 10.1177/1064804618788098. eprint: https://doi.org/10.1177/1064804618788098.

Woodruff, A., R. Rosenholtz, J. B. Morrison, A. Faulring, and P. Pirolli. (2002). "A Comparison of the Use of Text Summaries, Plain Thumbnails, and Enhanced Thumbnails for Web Search Tasks". *J. Am. Soc. Inf. Sci. Technol.* 53(2): 172–185. DOI: 10.1002/asi.10029.

Woods, W. A., R. M. Kaplan, and B. Nash-Webber. (1972). *The Lunar Sciences Natural Language Information System Final Report*. NASA.

Wu, C.-S., A. Madotto, E. Hosseini-Asl, C. Xiong, R. Socher, and P. Fung. (2019). "Transferable Multi-Domain State Generator for Task-Oriented Dialogue Systems". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 808–819. DOI: 10.18653/v1/P19-1078.

Wu, Z., B.-R. Lu, H. Hajishirzi, and M. Ostendorf. (2021). "DIALKI: Knowledge Identification in Conversational Systems through Dialogue-Document Contextualization". In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 1852–1863. DOI: 10.18653/v1/2021.emnlp-main.140.

Wu, Z., R. Parish, H. Cheng, S. Min, P. Ammanabrolu, M. Ostendorf, and H. Hajishirzi. (2022). "INSCIT: Information-Seeking Conversations with Mixed-Initiative Interactions". *CoRR*. abs/2207.00746.

Xiong, L., C. Xiong, Y. Li, K.-F. Tang, J. Liu, P. N. Bennett, J. Ahmed, and A. Overwijk. (2021). "Approximate Nearest Neighbor Negative Contrastive Learning for Dense Text Retrieval". In: *International Conference on Learning Representations*.

Xiong, W., X. Li, S. Iyer, J. Du, P. Lewis, W. Y. Wang, Y. Mehdad, W.-t. Yih, S. Riedel, D. Kiela, and B. Ouguz. (2020). "Answering Complex Open-Domain Questions with Multi-Hop Dense Retrieval". *ArXiv*. abs/2009.12756.

Xu, J., A. Szlam, and J. Weston. (2022). "Beyond Goldfish Memory: Long-Term Open-Domain Conversation". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 5180–5197. DOI: 10.18653/v1/2022.acl-long.356.

Xu, W. and A. Rudnicky. (2000). "Language modeling for dialog system." In: *INTERSPEECH*. 118–121.

Xu, X. and J. Lockwood. (2021). "What's going on in the chat flow? A move analysis of e-commerce customer service webchat exchange". *English for Specific Purposes*. 61: 84–96. DOI: https://doi.org/10.1016/j.esp.2020.09.002.

Yan, R., J. Li, and Z. Yu. (2022). "Deep Learning for Dialogue Systems: Chit-Chat and Beyond". *Found. Trends Inf. Retr.* 15(5): 417–589. URL: https://doi.org/10.1561/1500000083.

Yang, J.-H., S.-C. Lin, C.-J. Wang, J. Lin, and M.-F. Tsai. (2019). "Query and Answer Expansion from Conversation History". In: *TREC*.

Yang, L., M. Qiu, C. Qu, C. Chen, J. Guo, Y. Zhang, W. B. Croft, and H. Chen. (2020). "IART: Intent-Aware Response Ranking with Transformers in Information-Seeking Conversation Systems": 2592–2598. doi: 10.1145/3366423.3380011.

Yang, L., M. Qiu, C. Qu, J. Guo, Y. Zhang, W. B. Croft, J. Huang, and H. Chen. (2018a). "Response Ranking with Deep Matching Networks and External Knowledge in Information-seeking Conversation Systems". In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 245–254. doi: 10.1145/3209978.3210011.

Yang, L., H. Zamani, Y. Zhang, J. Guo, and W. B. Croft. (2017). "Neural Matching Models for Question Retrieval and Next Question Prediction in Conversation". *CoRR*. abs/1707.05409. url: http://arxiv.org/abs/1707.05409.

Yang, Y., Y. Gong, and X. Chen. (2018b). "Query Tracking for E-Commerce Conversational Search: A Machine Comprehension Perspective". In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 1755–1758. doi: 10.1145/3269206.3269326.

Yeh, Y.-T. and Y.-N. Chen. (2019). "FlowDelta: Modeling Flow Information Gain in Reasoning for Conversational Machine Comprehension". Nov.: 86–90. doi: 10.18653/v1/D19-5812.

Yenala, H., A. Jhanwar, and G. Chinnakotla M.K. an Goyal. (2018). "Deep learning for detecting inappropriate content in text". *Int J Data Sci Anal*. 6: 273–286. doi: 10.1007/s41060-017-0088-4.

Yu, D. and Z. Yu. (2021). "MIDAS: A Dialog Act Annotation Scheme for Open Domain HumanMachine Spoken Conversations". Apr.: 1103–1120. doi: 10.18653/v1/2021.eacl-main.94.

Yu, S., J. Liu, J. Yang, C. Xiong, P. Bennett, J. Gao, and Z. Liu. (2020). "Few-Shot Generative Conversational Query Rewriting". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1933–1936. doi: 10.1145/3397271.3401323.

Yu, S., Z. Liu, C. Xiong, T. Feng, and Z. Liu. (2021). "Few-Shot Conversational Dense Retrieval": 829–838. doi: 10.1145/3404835.3462856.

Zamani, H., M. Bendersky, X. Wang, and M. Zhang. (2017). “Situational Context for Ranking in Personal Search”. In: *Proceedings of the 26th International Conference on World Wide Web*. 1531–1540. DOI: 10.1145/3038912.3052648.

Zamani, H. and N. Craswell. (2020). “Macaw: An Extensible Conversational Information Seeking Platform”. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2193–2196. DOI: 10.1145/3397271.3401415.

Zamani, H. and W. B. Croft. (2020a). “Joint Modeling and Optimization of Search and Recommendation”. In: *Proceedings of the First International Conference on Design of Experimental Search and Information Retrieval Systems*. 36–41.

Zamani, H. and W. B. Croft. (2020b). “Learning a Joint Search and Recommendation Model from User-Item Interactions”. In: *Proceedings of the 13th International Conference on Web Search and Data Mining*. 717–725. DOI: 10.1145/3336191.3371818.

Zamani, H., S. T. Dumais, N. Craswell, P. N. Bennett, and G. Lueck. (2020a). “Generating Clarifying Questions for Information Retrieval”. In: *Proceedings of the 29th International Conference on World Wide Web*.

Zamani, H., G. Lueck, E. Chen, R. Quispe, F. Luu, and N. Craswell. (2020b). “MIMICS: A Large-Scale Data Collection for Search Clarification”. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 3189–3196. DOI: 10.1145/3340531.3412772.

Zamani, H., B. Mitra, E. Chen, G. Lueck, F. Diaz, P. N. Bennet, N. Craswell, and S. T. Dumais. (2020c). “Analyzing and Learning from User Interactions for Search Clarification”. In: *The 43rd International ACM SIGIR Conference on Research & Development in Information Retrieval*.

Zamora, J. (2017). “I’m Sorry, Dave, I’m Afraid I Can’t Do That: Chatbot Perception and Expectations”. In: *Proceedings of the 5th International Conference on Human Agent Interaction*. 253–260. DOI: 10.1145/3125739.3125766.

Zhai, C. (2016). "Towards a game-theoretic framework for text data retrieval". *IEEE Data Eng. Bull.* 39(3): 51–62. URL: http://sites.computer.org/debull/A16sept/p51.pdf.

Zhai, C. (2020). "Interactive Information Retrieval: Models, Algorithms, and Evaluation". In: 2444–2447. DOI: 10.1145/3397271.3401424.

Zhang, S. and K. Balog. (2020). "Evaluating Conversational Recommender Systems via User Simulation". In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1512–1520. URL: https://doi.org/10.1145/3394486.3403202.

Zhang, S., Z. Dai, K. Balog, and J. Callan. (2020a). "Summarizing and Exploring Tabular Data in Conversational Search". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1537–1540. DOI: 10.1145/3397271.3401205.

Zhang, S., M.-C. Wang, and K. Balog. (2022). "Analyzing and Simulating User Utterance Reformulation in Conversational Recommender Systems". In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 133–143. URL: https://doi.org/10.1145/3477495.3531936.

Zhang, S., H. Yang, and L. Singh. (2016). "Anonymizing Query Logs by Differential Privacy". In: *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 753–756.

Zhang, T., V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi. (2020b). "BERTScore: Evaluating Text Generation with BERT". In: *Proceedings of the 2020 International Conference on Learning Representations*.

Zhang, X., H. Xie, H. Li, and J. C.S. Lui. (2020c). "Conversational Contextual Bandit: Algorithm and Application". In: *Proceedings of The Web Conference 2020*. 662–672. DOI: 10.1145/3366423.3380148.

Zhang, Y. and C. Zhai. (2015). "Information Retrieval as Card Playing: A Formal Model for Optimizing Interactive Retrieval Interface". In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 685–694. DOI: 10.1145/2766462.2767761.

Zhang, Y., X. Chen, Q. Ai, L. Yang, and W. B. Croft. (2018). "Towards Conversational Search and Recommendation: System Ask, User Respond". In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 177–186. DOI: 10.1145/3269206.3271776.

Zhao, C., T. Yu, Z. Xie, and S. Li. (2022). "Knowledge-Aware Conversational Preference Elicitation with Bandit Feedback". In: *Proceedings of the ACM Web Conference 2022*. 483–492. DOI: 10.1145/3485447.3512152.

Zhao, T., K. Xie, and M. Eskenazi. (2019). "Rethinking Action Spaces for Reinforcement Learning in End-to-end Dialog Agents with Latent Variable Models". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 1208–1218. DOI: 10.18653/v1/N19-1123.

Zhou, X., L. Li, D. Dong, Y. Liu, Y. Chen, W. X. Zhao, D. Yu, and H. Wu. (2018). "Multi-Turn Response Selection for Chatbots with Deep Attention Matching Network". In: *ACL*.

Zue, V. W. and J. R. Glass. (2000). "Conversational interfaces: advances and challenges". *Proceedings of the IEEE*. 88(8): 1166–1180. DOI: 10.1109/5.880078.